68-77

REMARKS

Reconsideration of the present application is respectfully requested. Claims 1-7 are pending. Claims 1-4 and 7 have been amended. Claims 75-84 have been added. Support for the amendments is found in the claims as originally filed, and throughout the specification. No new matter has been added.

ELECTION

The election of Group I (claims 1-7) with traverse has been made final. Claims 8-74 are non-elected and withdrawn from consideration. Applicants reserve the right to pursue these claims in a continuing application.

Double Patenting Rejection

Claims 2-7 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-10 of U.S. Patent No. 6,235,972.

Applicant respectfully disagrees. The current application is a divisional application filed in response to the Restriction Requirement of U.S. Application Serial No. 09/426,557, now U.S. Patent No. 6,235,972. Claims 2-6 are not obvious in light of claims 1-10 of U.S. Patent No. 6,235,972. Claim 1 has been amended and claims an isolated polypeptide comprising an amino acid sequence having at least 80% sequence identity over the entire length of SEQ ID NO: 2 or 4, wherein the amino acid sequence encodes a polypeptide involved in nucleotide excision repair. Claims 3 and 4 have been amended and now depend from claim 1, therefore claims 3-7 are removed from the double patenting rejection. Claim 2 is directed to a recombinant expression cassette which expresses the polypeptide of claim 1 Therefore, the claim is defined by the polypeptide which was restricted out of U.S. Application Serial No. 09/413,574 as a separate invention. The claim is consonant with the polypeptide claim restricted, therefore it is respectfully requested that the double patenting rejection be withdrawn.

Rejection under 35 U.S.C §101 – Utility

Claims 1-7 are rejected under 35 U.S.C §101 because the claimed invention is not supported by either a specific and credible asserted utility or a well-established utility. The Examiner asserts for claim 1 (a) "It is unclear as to how this sequence would be useful, lacking its functional activity". The Examiner further states that claim 1 (b)-(d) lack a recitation of function.

Applicants respectfully disagree, page 2, lines 3-17 of the specification clearly details the well-established utility of Rad23 polypeptide involved in DNA nucleotide excision repair. In nucleotide excision repair, Rad23 is involved in transcription-coupled repair as well as over DNA repair. Rad23 has also been shown to interact with a number of proteins involved in DNA repair, transcription, proteolysis and cell cycle and has been implicated in ubiquitin mediated proteolysis and cell cycle regulation. The present invention proposes to use the well-established utility of Rad23 in order to modulate repair activity, and recombination, in order to improve gene targeting and transformation in plants (see page 2, lines 13-17) therefore establishing specific, substantial and credible utility for the present invention.

Claim 1 has been amended to delete parts a, b, and d, and further to recite 80% sequence identity over the entire length of the sequence, wherein the amino acid sequence encodes a polypeptide involved in nucleotide excision repair. Support for this amendment is found on page 2, lines 3-17, and page 50, lines 7-12 of the specification. New claims 75-84 have been added. New claims 75-77, which depend from claim 1, recite 85, 90, and 95% sequence identity respectively. Support for this amendment can be found on page 50, lines 7-12 of the specification. New claim 78, which depends on claim 1, recites the polypeptides of SEQ ID NOS: 2 and 4, and therefore finds support in original claim 1, part (b). Claim 1, parts (a) and (d) have been deleted and are not represented in the current amendment. As amended the claims now recite the functional activity.

It was well-known at the time of filing that Rad23 homologues are highly conserved throughout various eukaryotes from yeast to mammals to plants. The highly conserved amino acid motifs were also well known, for example, see Refs A1, A4, A7, A8, and A10 of the IDS submitted 11/12/2000 for Serial No. 09/413,574, now U.S. Patent 6,235,972. For example, multiple sequence alignments are presented in Appendix A, and further sequence analysis is presented in Appendix B. These analyses show the overall homology and conserved domains shared by the claimed sequences and other Rad23 sequences. Further, these analyses could be used to predict protein regions likely to be tolerant of amino acid substitutions and also indicate likely acceptable substitutions.

Appendix A presents two multiple sequence alignments performed using the PileUp program in the GCG sequence analysis suite. The first alignment shows SEQ ID NO: 2 aligned with Rad23 homologues from yeast (*S. cerevisiae*), human, rice and carrot. The second alignment contains SEQ ID NO: 4 aligned with the same sequences. The alignments indicate homology to Rad23 sequences across the entire length of SEQ ID NO: 2 and 4, and that SEQ ID NO: 2 and 4 show a high level of conservation with the ubiquitin-like (ubl), PAST, and ubiquitin-associated (uba) motifs of the Rad23 family, as seen by previous workers (for example, see Refs A1, A4, A7, A8, and A10 of the IDS submitted 11/12/00 for Serial No. 09/413,574, now U.S. Patent 6,235,972).

Appendix B presents sequence analyses of SEQ ID NO: 2 and SEQ ID NO: 4 performed by the BioScout suite of programs and includes the summary report, the 3D structure alignment to the UBA(2) domain of human Rad23A (HHR23A) crystal structure PDB 1F4I (Withers-Ward *et al.*, 2000 Biochemistry 39:14103-14112, enclosed, see also NMR structure Dieckmann *et al.* 1998 Nat. Struct. Biol. 5:1042-1047, enclosed), Pfam analyses results which indicate the presence of three Rad23 Pfam domains (SEQ ID NO: 2 e values = 7.8e –23, 9.5 e –12, and 4.5 e-06; SEQ ID NO: 4 e values = 2e –09, 2.7 e –11, and 2.3e –23), SEQ ID NO: 2 also shows an uncharacterized domain (e value 7.2), the alignments of the SEQ ID NO: 2 and 4

with the Pfam domains, and descriptions of the Pfam domains. Therefore, Applicant has shown the claimed sequences comprise extensive homology and conserved functional domains with known Rad23 sequences.

Support for polypeptides having a given percent sequence identity is taught in the specification as originally filed, for example, in SEQ ID NOS: 1-4; guidance on codon preferences (page 8, lines 5-25, and page 64, line 10 – page 65, line 6); codon degeneracy (page 6, line 23 – page 7, line 13, and page 27, line 31 – page 28, line 3); sequence comparisons (page 19, line 3 – page 24, line 24, and page 66, line 1 – page 69, line 3); conservative substitutions and variants (page 6, line 28 - page 8, line 4); polypeptide expression (page 42, line 15 – page 48, line 6, page 51, line 4 – page 54, line 16, and page 56, line 31 – page 57, line 12), isolation (page 57, line 14 – page 58, line 2) as well as the Examples.

Assays for Rad23 activity were known in the art at the time of filing. For example, nucleotide excision repair assays (*e.g.*, Mueller & Smerdon 1996, Ref. A3, Wang *et al.* 1997, Ref. A4, and Sugasawa *et al.* 1997, Ref. A6, in IDS submitted 11/12/99 for Serial No. 09/413,574, now U.S. Patent 6,235, 972), complementation of the yeast Rad23 mutants (*e.g.*, Watkins *et al.* 1993 Ref. A1, Mueller & Smerdon 1996, Ref. A3, and Sturm & Lienhard 1998 Ref. A10, in IDS submitted 11/12/99 for Serial No. 09/413,574, now U.S. Patent 6,235, 972). Also available were yeast two-hybrid screens, gene knockouts, UV sensitivity, binding assays, co-purification, gels, blots, co-immunoprecipitation, and immunolocalization. Therefore, one of skill in the art could readily identify macromolecules having Rad23 activity.

Applicants believe that the claimed invention has a well-established utility for which they have proposed specific, substantial and credible uses. Therefore, it is respectfully requested that the rejection of claims 1-7 under 35 U.S.C §101 should be withdrawn and not applied to new claims 75-84.

Rejections under 35 U.S.C. §112, 1st Paragraph, Enablement:

Claims 1-7 are rejected under 35 U.S.C. §112, first paragraph, because the specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention commensurate in scope with these claims. Specifically, since the claimed invention lacks utility for the reasons set forth in the rejection under 35 U.S.C. §101, one skilled in the art would not know how to use the invention.

Applicants believe they have properly addressed the rejection under 35 U.S.C. §101 by amendment and argument and therefore respectfully request that the corresponding rejection of claims 1-7 under 35 U.S.C. §112, first paragraph for lack of enablement be withdrawn.

Claims 1-7 are further rejected under 35 U.S.C. §112, first paragraph, because the specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention commensurate in scope with these claims.

The Examiner asserts that the specification does not provide enablement for an isolated polypeptide comprising at least 20 contiguous amino acids of SEQ ID NO: 2 or 4, a polypeptide of SEQ ID NO: 2 or 4, a polypeptide having at least 70% sequence identity to SEQ ID NO: 2 or 4, or to a polypeptide encoded by a polynucleotide that will hybridize under stringent conditions to SEQ ID NO: 1 or 3. The Examiner asserts that the specification only teaches that a RAD23 gene of yeast was known and that the Applicant believes that the disclosed sequences are of a maize RAD23. The Examiner states that there is no evidence that SEQ ID NO: 2 or 4 have similar structural features to the yeast RAD23 gene, nor is there evidence of functional activity for the claimed polypeptides. The Examiner states that the specification fails to provide guidance regarding how one skilled in the art would evaluate SEQ ID NO: 2 or 4 for a specific activity.

Applicants respectfully disagree. Applicants disclose two polynucleotide sequences, each of which encode a full-length polypeptide having overall homology to known Rad23 sequences and further containing conserved motifs seen in the known Rad23 sequences (see above). It is respectfully submitted that 35 U.S.C. §112 does not require a working example. Further, case law does not require that an application provide a working example. It is only required that the application teach the person skilled in the art how to make and use the invention. The present application meets that requirement.

Similarly, the disclosure of SEQ ID NOS: 1-4, the conserved domains and motifs known in the art as shown in Appendix A & B and the IDS filed 11/12/99, the sequence identity and similarity to other known Rad23 sequences, the guidance on sequence analyses, comparison, and identity, the guidance on codon degeneracy, silent variants, and preferences, the guidance on conservative amino acid substitutions, and the ready availability in the art for assays for Rad23 (all as noted above) show the specification coupled with the knowledge in the art enables a person in the art to make and use sequences having 80% sequence identity to SEQ ID NO: 2 or 4 involved in nucleotide excision repair.

The present invention discloses how to make and use the sequences of the invention, as discussed above. The question of experimentation is a matter of degree. The fact that some experimentation is necessary does not preclude enablement; what is required is the amount of experimentation must not be unduly extensive. *PPG Inc. v. Guardian Industries Corp.* (37 USPQ 1218, 1623, (Fed. Cir. 1996). The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the invention claimed. *Ex parte Jackson*, 217 USPQ 804, 807 (1982 PTOBA).

The Examiner asserts that sequence homology is not sufficient to predict function, citing Doerks *et al.* (*TIG* 14(6):248-250 1998), Smith *et al.* (*Nature Biotech.* 15:1222-1223 1997), Brenner (*TIG* 15(4):132-133 1999), and Bork (*TIG* 12(10):425-327 1996).

Applicant respectfully disagrees. These references all point to the potential accumulation and propagation of errors from automatic computer annotation of high throughput sequencing. This kind of sequencing and annotation was frequently a one-pass sequencing and annotation by top local alignment hit. These do not state that analyses of sequence homology are not predictive of function, these point to the problem of automatic annotation based on one top hit in a database. Doerks *et al.* actually try to predict the function of uncharacterized protein families (UPFs) through sequence analyses alone (page 248, column 1-3). Smith points out that the problems do not lie in the homology search algorithms, but in minor database annotation inconsistencies (page 1222, col. 1, 3rd paragraph – col. 2, 1st paragraph). Brenner's statement that ".. then most homologs must have different molecular and cellular functions" is prefaced by "**if** there are only about a 1000 major superfamilies in nature" (emphasis added). Brenner examined the annotations by three groups for the complete genome of *M. genitalium*, and estimated the minimum error rate of annotations to be 8%, therefore the correct annotation rate would be approximately 92%. Bork states "we wish to point out that sequence databases are the most useful tool in sequence analysis and the question should be how can one further improve their value" (page 472, col. 1, 1st full paragraph). Taken as a whole, these references warn of the potential error of relying on one automated annotation to identify a gene homologue, but do not state that homology is NOT indicative of function.

Examiner cites Van de Loo et al. (PNAS 92:6743-6747 1995) as an example that gene activity cannot be determined merely by similarity of sequences. Examiner states that the reference teaches that a change in only four amino acids will convert a desaturase gene to a hydroxylase gene.

Applicant respectfully disagrees. It was the similarity of the activities, cofactors and substrates that suggested to van de Loo *et al.* that there might be sequence similarity as well (page 6743, col. 1-2). The hydroxylase could not be purified by biochemical means, therefore van de Loo *et al.* used differential screening and sequencing to identify seed-specific clones that were similar to desaturase. A hydroxylase cDNA clone was identified and confirmed based on its similarity to a desaturase. Applicant found no teaching of amino acid changes to convert a desaturase to a hydroxylase in this reference. Applicant believes the Examiner may be referring to Broun *et al.* (Science 1998 282:131-133) in which very specific limited amino acid changes which resulted in alteration of the experimental protein's catalytic activity. Broun discloses modifications to a oleate 12-desaturase sequence which resulted in the protein having hydroxylase activity. Broun *et al.* actually note the high sequence similarity between the oleate 12-desaturase and oleate hydroxylase and use this to identify seven residues conserved in desaturases and to target them for modification and activity (see page 131, column 2 – column 3). Broun *et al.* actually use the sequence similarity of the desaturase and the hydroxylase to predict which residues to change to alter the activity of the desaturase.

Applicants submit that no more than routine experimentation is required. This may be accomplished by the examples and methods within the present application and within the technical, scientific, skill in the art. The disclosure of SEQ ID NOS: 1-4, the knowledge in the art of other Rad23 sequences and the conserved regions, the guidance on conserved motifs and amino acids, and sequence comparisons as discussed above, and the ready availability of routine Rad23 assays enabled one of skill in the art to make and use the claimed invention. Therefore it is respectfully requested that the rejection of claims 1-7 under 35 U.S.C §112, first paragraph be withdrawn, and not applied to new claims 75-84.

R jections under 35 U.S.C. §112, 1st Paragraph, Written Description:

Claims 1-7 are rejected under 35 U.S.C. §112, first paragraph, as containing subject matter not sufficiently described in the specification to indicate the inventor(s) had possession of the invention.

The Examiner asserts that the Applicant has not identified structural features necessary to describe features of SEQ ID NO: 2 or 4 which distinguish them from any polypeptide, since 30% lack of sequence identity could fall in a region essential for protein activity. The Examiner states that a genus is claimed, but that there is no description of the structural feature that defines the genus, citing *University of California v. Eli Lilly*, 119 F.3d 1559, 43 USPQ 2d 1398 (Fed. Cir. 1997). The Examiner concludes that given the lack of written description in the specification regarding the structural and physical characteristics of the claimed compositions one skilled in the art would not have been in possession of the genus claimed at the time this application was filed.

The Examiner is reminded that every species encompassed by the claimed invention need not be disclosed in the specification to satisfy the written description requirement of 35 U.S.C. § 112, first paragraph. Utter v. Hiraga, 845 F.2d 993, 6 USPQ2d 1709 (Fed. Cir. 1988). In fact, the description of a claimed genus can be by structure, formula, chemical name, or physical properties. See Ex parte Maizel, 27 USPQ2d 1662, 1669 (B.P.A.I.1992), (citing Amgen v. Chugai, 927 F.2d 1200, 1206 (Fed. Cir. 1991)).

Claim 1 has been amended to recite that the sequence shares 80% sequence identity to the sequence of SEQ ID NO: 2 or 4, and further recites that the sequence encodes a polypeptide involved in nucleotide excision repair. The recitation of at least 80% sequence identity is a very predictable structure of the sequences encompassed by the claimed invention. Further, Applicants have provide two isolated sequences, SEQ ID NOS: 2 and 4. Contrary to yeast, the presence of two Rad23 genes was also seen in other higher eukaryotes such as human, mouse, and carrot. (Refs. A4, A6, and A10 in IDS submitted 11/12/99 for Serial No. 09/413,574,

now U.S. Patent 6,235, 972). Further, Sturm *et al.* (Ref A10) showed that both carrot Rad23 homologues complemented a yeast *rad23* mutant. The description of a representative number of species does not require the description to be of such specificity that it would provide individual support for each species that the genus embraces. 66 Fed. Reg. 1099, 1106 (2000). Satisfactory disclosure of a "representative number" depends on whether one of skill in the art would recognize that the Applicants were in possession of the necessary common attributes or features of the elements possessed by the members of the genus in view of the species disclosed. 66 Fed. Reg. 1099, 1106 (2000). Applicants submit that the knowledge and level of skill in the art would allow a person of ordinary skill to envision the claimed invention, *i.e.*, a sequence having at least 80% sequence identity to the sequence set forth in SEQ ID NO: 2 or 4. The recitation of a predictable structure of at least 80% sequence identity to SEQ ID NO: 2 or 4 is sufficient to satisfy the written description requirement.

In addition, an Applicant may rely upon functional characteristics in the description, provided there is a correlation between the function and structure of the claimed invention. Claim 1 recite that the claimed sequences are, or encode, a polypeptide involved in nucleotide excision repair thereby providing a functional characterization of the sequences claimed in the genus.

Example 14 of the Revised Interim Written Description Guidelines is directed to a generic claim: a protein having at least 95% sequence identity to the sequence of SEQ ID NO:3, wherein the sequence catalyzes the reaction A → B. The Training Materials concludes that the generic claim of Example 14 is sufficiently described under § 112, first paragraph, because 1) "the single sequence disclosed in SEQ ID NO:3 is representative of the genus" and 2) the claim recites a limitation requiring the compound to catalyze the reaction from A → B. The Guidelines conclude that one of skill in art would recognize that the Applicants were in possession of the necessary common attributes possessed by the members of the genus.

Following the analysis of Example 14, Applicants submit that claim 1 satisfies the written description requirements of § 112, first paragraph. Specifically, the claims of the present invention encompass sequences having at least 80% sequence identity to the sequence of SEQ ID NO: 2 or 4, wherein the claimed sequences encode a polypeptide involved in nucleotide excision repair. As in Example 14, the specification discloses the amino acid sequences of SEQ ID NOS: 2 and 4, and the amended claims recite a limitation requiring the sequence to have a specific function.

Consequently, contrary to the Examiner's conclusion, the sequences encompassed by the genus of claim 1 are defined by relevant identifying physical and chemical properties. In fact, the common attributes or features of the elements possessed by the members of the genus is that they are Rad23 sequences and share at least 80% sequence identity at the amino acid level to the disclosed polypeptide sequences of SEQ ID NOS: 2 or 4. The necessary common features of the claimed genus are clear.

Applicants have disclosed two isolated polynucleotide sequences and the encoded Rad23 polypeptides in SEQ ID NOS: 1-4, and have provided further guidance on sequence isolation, analysis, and identification, codon degeneracy, conserved protein domains and motifs, as well as conservative amino acid substitutions. Applicants clearly had possession of sequences having at least 80% sequence identity to SEQ ID NO: 2 or 4, therefore the rejection of claims 1-7 under 35 U.S.C. §112, first paragraph written description should be withdrawn.

Rejections under 35 U.S.C. §112, 2nd Paragraph:

Claims 1-7 are rejected under 35 U.S.C. §112, second paragraph, as being indefinite.

The Examiner asserts that the term "member" renders claim 1 indefinite. Also, the Examiner asserts that the terms "selectively hybridize" and "stringent hybridization conditions" in claim 1 (d) are unclear.

The Applicant respectfully disagrees that recitation of a "member selected from the group" in the preamble makes claim 1 indefinite since this is common Markush group claim format and parts (a) - (d) of the claim clearly recited a "polypeptide". However, in order to expedite prosecution, claim 1 has been amended to delete parts (a), (c) and (d), and no longer recites "a member". Claim 1 part (d) has been deleted and not represented Therefore, the rejection of claims 1-7 under 35 U.S.C. §112, second paragraph is obviated.

Applicants have properly addressed by argument and amendment the grounds for the rejection of claims 1-7 under 35 U.S.C. §112, second paragraph and respectfully request that the rejection be withdrawn.

CONCLUSION

In light of the foregoing remarks and amendments, it is believed that claims 1-7, and 75-84 are in condition for allowance. Withdrawal of the outstanding rejections and allowance of all of the remaining claims is respectfully requested.

Respectfully submitted,

Virginia Dress
Agent for Applicant(s)
Registration No. 48,243

PIONEER HI-BRED INTERNATIONAL, INC.
Corporate Intellectual Property
7100 N.W. 62nd Avenue
P.O. Box 1000
Johnston, Iowa 50131-1000
Phone: (515) 270-4192
Facsimile: (515) 334-6883

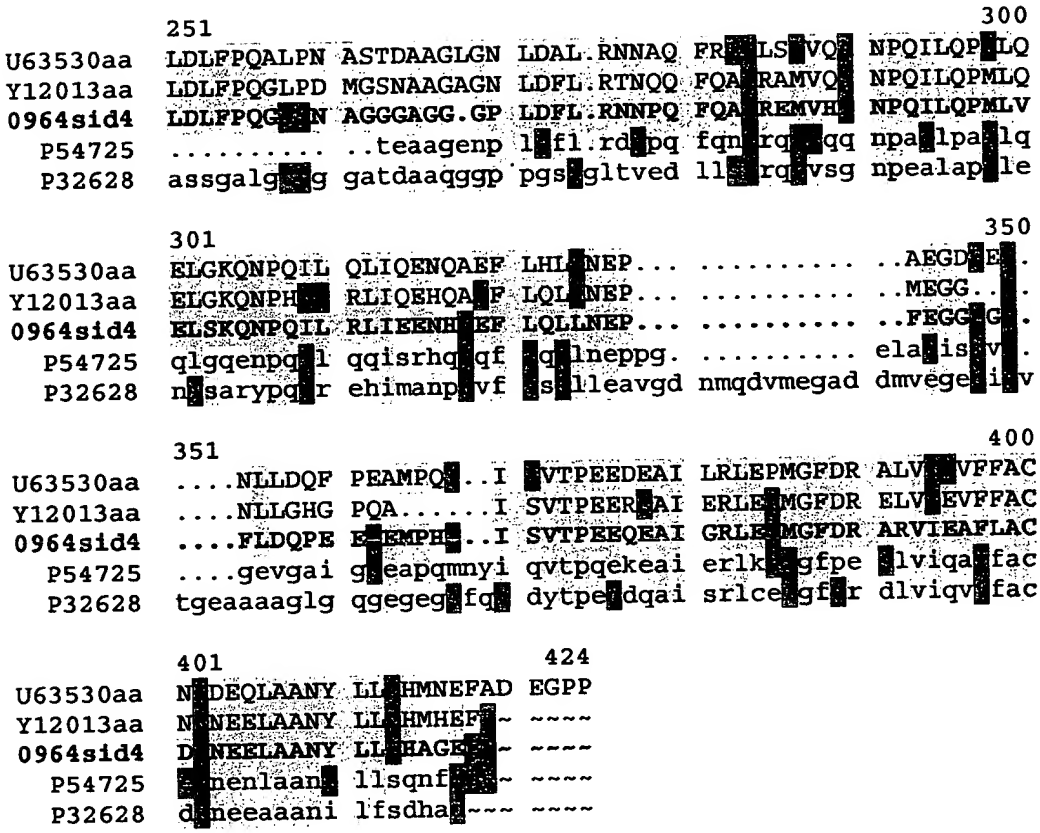
U.S. Serial No. 09/805,550
Group Art Unit: 1638

APPENDIX A

```
!!AA_MULTIPLE_ALIGNMENT 1.0
PileUp of: @/tmp/54762539.list

 Symbol comparison table: genrundata:blosum62.cmp  CompCheck: 1102

                    GapWeight: 8  GapLengthWeight: 2

 0964DP32628_pileup_54762.txt  MSF: 433  Type: P  March 31, 2003 14:25  Check:
7851 ..

Y12013aa     Protein encoded by GenBank Y12013, Carrot Rad23
0964sid2     SEQ ID NO: 2, Maize Rad23
U63530aa     Protein encoded by GenBank U63530, Rice Rad23
P54725       SwissProt protein deposit P54725, human Rad23A
P32628       SwissProt protein deposit P32628, S. cerevisaie Rad23

Residues IDENTICAL to 0964sid2.

                1                                                    50
     Y12013aa   ~~MK YVKTL KG FEIQ N PDDSVADVKR SIET QGA V YPAAQQMLIY
     0964sid2   ~~MKLNVKTL KG FEIEAS PDASVADVKR IIETTQGQS  YRA QQMLIY
     U63530aa   ~~MK SVKTL KG TFQIE  SAQKVADVKR IIETTQGQHI YPA QQMLIH
       P54725   ma t t ktl qqqtfkir  pde vkv ke kie ekgrd  pvagqkliy
       P32628   ~m sltfknf kkekvp d  psn l tkt k  qsisc .eesqikliy

                51                                                   100
     Y12013aa   QGK LKDGTT LLENNVAENS F VIMLSK K PSGEGST  TAAAPK PQT
     0964sid2   QGKILKD TT L SNGVAENS FLVIMLSK K ASSSGAST  TAKAPA L Q
     U63530aa   QGK LKD TT L ENKVLENS FLVIMLRQGK GSSSSAP  KAP NQ PPT
       P54725   agkilsd vp  rdyr dekn f v m k k agqgt ppe  p aape s
       P32628   sgk lqdskt  secg k gd q vfm sqkk  k.....k v eppi pes

                101                                                  150
     Y12013aa   SAPPSVPAP  V QPP S LP VP PSPAP.. .APA APIP  AAVGS ANVY
     0964sid2   PAAPVAPA   V RTPTQA.P V  AE APP  VQPQAAPAA  VA  D ADVY
     U63530aa   Q VPAAPA Q APVAP T VP V V AP PT  T...A PAP  VAV S ADNY
       P54725   t fppap g  hpppaare dk pseesap ttspe v g  vp gs gre
       P32628   a tpgren .  .....teasp s das apa  tapeg qpqe eq tterte

                151                                                  200
     Y12013aa   DSAASLLVAG SNLEG IQQI LDMGGGTW R DTV RI RAA  NNPERA  Y
     0964sid2   SQAASNLVFG NNLEQTIQQI LDMGGGTWER DTVVRALRAA YNNPERA  Y
     U63530aa   GQA SNLVAG SNLEATIQSI LDMGGGIW R DIV HALSAA  NNPERA  Y
       P54725   edaastlvtg seyetm tei  smg...er  rvvaalra  nnphra  y
       P32628   sa pgfvvg ternetieri  mg...qr  everalraa  nnp ra  y

                201                                                  250
     Y12013aa   LYSGIPE AE APPVAP PP  G.....QAAN PLDQPPAA.. AQP PASAGP
     0964sid2   LYSGIPEN E AQPVAR PAA GQQ NQQAAS PA.QPAVALP VQPSPASAGP
     U63530aa   LYSG PE   I.PV..PPP  IQP NPTQAS QATQPAA... ..PSILS GP
       P54725   ll gip.... ......gspe pehgsvqe q v eqpa.... ..........
       P32628   llmgipen r qpepqqqtaa aaeqpstaa  taeqpaeddl f..qa qgg
```

```
           251                                                      300
Y12013aa   NANPLDLFPQ G PDMGSN..  ...AAG GNL DFLRTNQQFQ ALRA VQ NP
0964sid2   NANPLNLFPQ G PGGSNPG  VVPGAG GAL DALR LPQFQ ALLQLVQANP
U63530aa   NASPLDLFPQ A PNAS D..  ...AAGLGNL DALR NAQFR LLLSLVQANP
  P54725   .......... .......... .teaagenpl flrdqpqfq n rq qqnp
  P32628   nassgalgtt gga da gg  ppgsigltve dll.......  lrq vsgnp

           301                                                      350
Y12013aa   QILQPMLQEL GKQNPH RL  IQEHQA FLQ L NEPM.... ..........
0964sid2   QILQPMLQEL GKQNPQILRL  IQENQAEFLR L NE P.... ..........
U63530aa   QILQP LQEL GKQNPQILQL  IQENQAEFLH L NEPA.... ..........
  P54725   a lpa lqql gqenpq lqq  isrhqeqf q  neppg... .........el
  P32628   ealap len  sarypq reh  imanpevf s  le vgdnm qdvmegaddm

           351                                                      400
Y12013aa   .EGGE..N L GH....GPQ.  ....  ..  V TPEER AIER LEAMGFDREL
0964sid2   .EGGPGGN L GQLAAA PQ.  ....T..  V TPEEREAIQR LEGMGFNREL
U63530aa   .EGDDEEN L DQFPEA PQ.  ....T..  V TPEEDEAILR LEPMGFDRAL
  P54725   a isdvege  gaigeeapq.  ....mny qv tpqe eaier lka gfpesl
  P32628   vegedievtg eaaaag gqg  egeg fq dy tpeddqaisr lce gfer l

           401                                  433
Y12013aa   VLEVFFACNK NEEL ANYLL .DHMHEFE~~ ~~~
0964sid2   VLEVFFACNK DEEL ANYLL .DHGHEFDDQ QQ~
U63530aa   VL VFFACNK DEQL ANYLL .DHMNEFADE GPP
  P54725   v qa facek nenl an ll .s..qnfdde ~~~
  P32628   v qv facdk neea anilf sdhad~~~~~ ~~~
```

```
!!AA_MULTIPLE_ALIGNMENT 1.0
PileUp of: @/tmp/54847313.list

 Symbol comparison table: genrundata:blosum62.cmp  CompCheck: 1102

                  GapWeight: 8   GapLengthWeight: 2

 0964DP32628_pileup_54847.txt  MSF: 424  Type: P  March 31, 2003 14:40  Check:
7143 ..

 U63530aa     Protein encoded by GenBank U63530, Rice Rad23
 Y12013aa     Protein encoded by GenBank Y12013, Carrot Rad23
 0964sid4     SEQ ID NO: 4, Maize Rad23
 P54725       SwissProt protein deposit P54725, Human Rad23A
 P32628       SwissProt protein deposit P32628, S. cerevisaie Rad23

Residues IDENTICAL to 0964sid4.

                1                                                    50
U63530aa  ~~MK VKTL KG TFQIEVD SAQK ADVK  IIETTQGQHI YPAEQQ LIH
Y12013aa  ~~MK YVKTL KG QFEIQV  PDD  ADVK  SIETAQGAAV YPAAQQ LI
0964sid4  ~~MKLTVKTL KG HFEIRV  PNDTI AVK  NIEEIQG D  YPWGQQ LI
  P54725  ma t t ktl qqqtfkir e pd t kv ke kieaekg d  pvagqkli
  P32628  ~m sltfknf kkekvp d e psnti etkt k aqsiscee ...sqikli

               51                                                   100
U63530aa  GKVLKD T L ENKVLENS FLV MLRQGK GSSS APATS K PSN PPT
Y12013aa  GKVLKDG T LLENNVAENS F V MLSKSK PSGEGSTTS  AAPK PQT
0964sid4  GKVLKD T L ENKVNEDG FLVVMLSKGK SGSTG..TS  SQH P T
  P54725  agk lsd vp  r y dekn f vvm ktk  gqgtsappe  spt ape
  P32628  sgkvlqdskt  secg kggd q vfm sqkk  k.....tk v eppi pe

               101                                                  150
U63530aa  QTVPAAPAS. QAPVAPA .. .TVP TVSAP TPTA A PAP AVAVSSEADN
Y12013aa  SAPPS PAP. AVSQPPA .. .TLP P PSP APAPATAPIP SAAVGSEANV
0964sid4  RQ PP EAPQ QAPQPPVA.. ...P T SQP EGLPAQAP.. ........NT
  P54725  tsfppap sg mghpppaare dkspseesap ttsp.e vsg svpssgssgr
  P32628  at pgrenst eas..ps da saapa. ap egsqpqeeqt attertes..

               151                                                  200
U63530aa  YG A SNL  GSN  TI S  EMGGGIWD  DIVLHALSA A NNPERAVE
Y12013aa  YDSAASLL  GSN  GAI Q  MGGG WD  DTVIRI RA A NNPERAVE
0964sid4  HD AASNL  GRNV T I Q  MEMGGG WD  D VQRALRA AYNNPERAVE
  P54725  e daastl  gsey t te  msmg...  vvaalra ynnphrave
  P32628  ..a pgf v gternetier memg...q  everalra annp rave

               201                                                  250
U63530aa  YLYSG PEQM IP..VP.PP SIQPANPTQA SQATQP AP  ILSSGPN P
Y12013aa  YLYSGIPEQA EAPPVAPSPP SGQAANPLDQ PPAAAQPAPA ..SAGPN NP
0964sid4  YLYSGIPVTA EI....AVPI GGQGAN TDR ....AP G A GLSGIPN P
  P54725  yll gip...  ....gspepe hgs....vqe sqvseqpa.. ..........
  P32628  yllmgipenl rqpepqqqta aaaeqp taa ttaeqp e d lf qaa ggn
```



U.S. Serial No. 09/805,550
Group Art Unit: 1638

APPENDIX B

bioSCOUT®

Projects | Analyses | Admin | Alert | SRS | Tutorial | Help

Analysis Browser: Level Up

Report for 0964rad23.sid2 (Protein) Update

Description 0964_rad23_sid2 Edit

Function

At5g38470/At5g38470

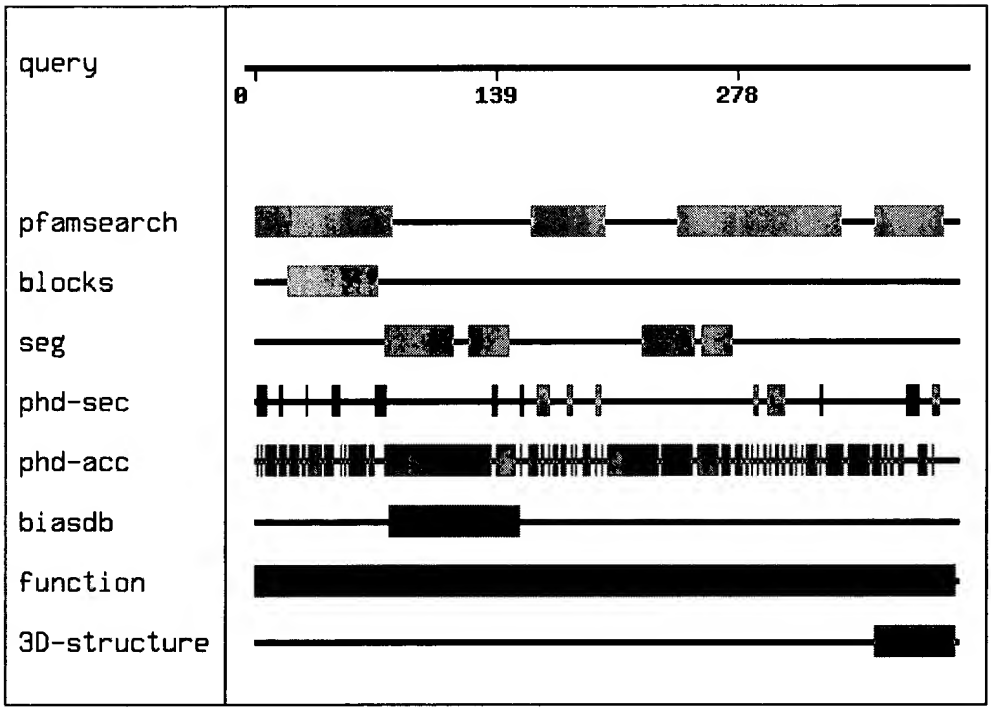
Direct assignment of functionality by homology to tremblnew|AY081835|AY081835_1

in region **1** to **403** for overall length of **228** (99% of query, 176% of hit, see the alignment).

Functional class Replication

Extracted keywords DNA damage, Nuclear protein, DNA repair

Features Summary



Homologies

All BLAST hits

Protein **68 clear** homologs All protein BLAST hits

	ESTs	**206** homologs	All EST BLAST hits
	Patents	**103** homologs	All patent hits
General	**Gene name**		
	Molecular weight	42.64 kD	
	Sequence length	405	
	Isoelectric point	4.40	
	Predicted cellular localisation (PHD and PreLoc)	extracellular (86.8 %)	
3D Structure	**3D structure inferred by clear homology from residues 357 to 403 in 1F4I-A**		
	View	alignment	
	pdb|1F4I|1F4I-A	structure	
Expression	**Expression of this gene is reported for**		
	Organ Category	other species	
	Tissue Classification	normal tissue	
Phylogeny	**Distribution**	32 species extracted from 203 homologous sequences.	Species
	Taxa	Chordata, Eukaryotae, Fungi, Planta	
	Model organisms	*Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Homo sapiens, Mus musculus, Saccharomyces cerevisiae*	
Features	**Low complexity region**	from **75** to **114,** from **123** to **146,** from **223** to **253,** from **257** to **275** detected by [seg]	
	A-rich region	from **77** to **152** detected by [biasdb]	
	No significant hits detected by	[Coils] [Phd-tm]	
Patterns	**UBA/TS-N domain region**	from residue **357** to **396.** Source: [pfamsearch] . Quality: (E=4.5e-06)	
	Uncharacterised protein family (UPF0184) - region	from residue **244** to **337.** Source: [pfamsearch]	

UBA/TS-N domain region	from residue **159** to **201.** Source: [pfamsearch] . Quality: (E=9.5e-12)
Ubiquitin family region	from residue **1** to **78.** Source: [pfamsearch] . Quality: (E=7.8e-23)
No significant hits found in	[prosite database] [blocks database]

Comment No comment section. Edit

Completed Tasks Interactiv

Start Time	User	Comment	Output	
03.03.2003, 15:36:58	dressvm		bioSCOUT_default	details...

Permissions Edit

Alert Jobs New

Please report problems and feedback concerning bioSCOUT through the support interface.

Alignment: 0964rad23.sid2 - pdb|1F4I|1F4I-A

BLASTP - alignment of 0964rad23.sid2 against pdb|1F4I|1F4I-A

uv excision repair protein protein rad23 homolog afragment: c-terminal uba domain;Mutant

- This hit is scoring at : 9e-06 (expectation value)
- Alignment length (overlap) : 47
- Identities : 48 %
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)
- Database searched : nrdb

```
Q:    357 EEREAIQRLEGMGFNRELVLEVFFACNKDEELTANYLLDHGHEFDDQ     403
          :E:EAI:RL:.:GF...LV:::FAC.K:E.L.AN:LL  ...FDD:
H:      1 QEKEAIERLKALGFEESLVIQAYFACEKNENLAANFLL--SQNFDDE      45
```

Legend of Alignment

```
:  positive score

.  score between -2 and 0
```

Please report problems and feedback concerning bioSCOUT through the support interface.

Biochemical and Structural Analysis of the Interaction between the UBA(2) Domain of the DNA Repair Protein HHR23A and HIV-1 Vpr†

Elizabeth S. Withers-Ward,‡,⊥ Thomas D. Mueller,§,⊥ Irvin S. Y. Chen,*,‡ and Juli Feigon*,§

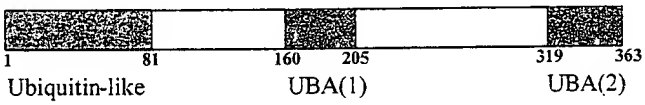
Departments of Microbiology, Immunology and Medicine, and UCLA AIDS Institute and Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90095

Received July 24, 2000; Revised Manuscript Received September 27, 2000

ABSTRACT: The DNA repair protein HHR23A is a highly conserved protein that functions in nucleotide excision repair. HHR23A contains two ubiquitin associated domains (UBA) that are conserved in a number of proteins with diverse functions involved in ubiquitination, UV excision repair, and signaling pathways via protein kinases. The cellular binding partners of UBA domains remain unclear; however, we previously found that the HHR23A UBA(2) domain interacts specifically with the HIV-1 Vpr protein. Analysis of the low resolution solution structure of HHR23A UBA(2) revealed a hydrophobic loop region of the UBA(2) domain that we predicted was the interface for protein/protein interactions. Here we present results of in vitro binding studies that demonstrate the requirement of this hydrophobic loop region for interaction with human immunodeficiency virus (HIV-1) Vpr. A single point mutation of the Pro at residue 333 to a Glu totally abolishes the binding of HIV-1 Vpr to UBA(2). High resolution NMR structures of the binding deficient UBA(2) mutant P333E as well as of the wild-type UBA(2) domain were determined to compare the effect of this mutation on the structure. Small but significant differences are observed only locally at the site of the mutation. The biochemical and structural analysis confirms the function of the HHR23A UBA(2) GFP-loop as the protein/protein interacting domain.

HHR23A is one of the human homologues of the yeast Rad23 and Rhp23 proteins that functions in nucleotide excision repair (*1*−*3*). The biochemical function of HHR23A is poorly understood. In yeast, *rad23* mutants are sensitive to UV light (*4*). Rad23 promotes the assembly of a large multiprotein complex involved in nucleotide excision repair (*5*). In *Saccharomyces cerevisiae*, Rad23 is complexed with Rad4. Similarly, both of the human homologues termed HHR23A and HHR23B interact with the human ortholog, the Xeroderma pigmentosum group C (XPC) protein and promote its activity in in vitro DNA repair assays (*3*, *6*−*9*). In vitro studies have shown that the Rad23/Rad4 complex preferentially binds to damaged DNA (*10*, *11*). Studies in yeast indicate that Rad23 is required for both transcription-coupled DNA repair and efficient global genome repair (*12*).

All of the Rad23 homologues share a common domain structure, including a ubiquitin-like domain at the N-terminus and two copies of a highly conserved domain termed the ubiquitin associated domain (UBA) located in the middle [UBA(1)] and at the C-terminus [UBA(2)] of the protein (*6*) (Figure 1). The N-terminal ubiquitin-like (UBL) domain interacts with the 26S proteasome, as does that of HHR23A and B, suggesting a role for Rad23 in protein degradation via the ubiquitin−proteasome pathway (*13*−*15*). In yeast, the UBL domain is required for function (*2*). UBA domains are approximately 45 amino acids long and are found in many proteins including those involved in the ubiquitination pathway, UV excision repair, and cell signaling (*16*). The C-terminal UBA domain of HHR23A has very recently been shown to interact with the human DNA base excision repair protein 3-methyladenine DNA glycosylase (MPG) (*17*).



HHR23A

FIGURE 1: Schematic representation of the domain structure of HHR23A. The shaded boxes indicate the domains that have been identified by sequence analysis and include an N-terminal ubiquitin-like domain and two copies of highly conserved ubiquitin associated domain (UBA) which are designated UBA(1) and UBA(2).

We previously found that the human immunodeficiency virus (HIV-1) Vpr protein binds specifically to the UBA(2) domain of HHR23A (*18*). The Vpr protein is a 96-amino acid gene product of HIV-1 that is incorporated into virions (*19*). Vpr plays a role in several aspects of the viral life cycle including nuclear import of the preintegration complex, transactivation of transcription, induction of cell cycle arrest at the G2/M checkpoint, and subsequent apoptosis (*20*−*24*). These functions are mediated by distinct domains of Vpr. Residues within both the N- and C-termini are implicated in mediating cell cycle arrest (*25*, *26*). Selection for maintenance of wild-type Vpr occurs in chimpanzees and

† This work was supported by NIH Grant AI43190 to I.S.Y.C. and J.F.

* To whom correspondence should be addressed. E-mail: feigon@mbi.ucla.edu; phone: (310) 206-6922; fax: (310) 825-0982.

‡ Departments of Microbiology, Immunology and Medicine, and UCLA AIDS Institute.

§ Department of Chemistry and Biochemistry and the Molecular Biology Institute.

⊥ These authors contributed equally to this work.

10.1021/bi0017071 CCC: $19.00
Published on Web 10/27/2000

humans infected with HIV-1, supporting the idea that Vpr confers an important selective advantage to the virus (*27*).

Vpr-induced cell-cycle arrest and apoptosis of HIV-1 infected cells would have a profound effect on T-cells, the primary target of HIV-1, contributing to the immune deregulation associated with AIDS (*28*). HIV-1 infected $CD4^+$ T-cells would fail to undergo activation and clonal expansion. Vpr-mediated apoptosis would contribute to depletion of the $CD4^+$ cells that is seen in HIV-infected individuals. Therefore, Vpr may be an important viral protein to target in the development of therapeutic reagents for treatment of HIV-1-infected individuals. In addition to finding that Vpr interacts specifically with the UBA(2) domain of HHR23A, we previously demonstrated that overexpression of HHR23A or a fragment encompassing the UBA(2) domain would partially alleviate cell cycle arrest mediated by Vpr (*18*). This suggests a significant functional role for the interactions between Vpr and HHR23A. We hypothesized that Vpr may mediate cell cycle arrest and/or apoptosis by binding to the UBA(2) domain of HHR23A, thereby leading to deregulation of normal cellular processes involved in the control and signaling of DNA repair. Thus, an understanding of the structural requirements for interaction between Vpr and UBA(2) may not only provide insight into the role of HHR23A in Vpr function but may also lead to strategies for the development of therapeutic reagents that would interfere with Vpr function.

We have previously reported a low resolution structure of UBA(2) of HHR23A determined by homonuclear NMR spectroscopy (*29*). The structural motif of the UBA(2) domain is composed of three α-helices folded around a hydrophobic core. Analysis of the NMR structure revealed an unusually large hydrophobic area of the protein that we predicted to be an interface for protein/protein interactions and might be the Vpr-interacting region of the HHR23A UBA(2) domain. This hydrophobic area of the UBA(2) domain is in the vicinity of the GFP-loop ($_{331}$GFP$_{333}$) which forms a sharp turn between α-helices 1 and 2. To further characterize the amino acids that are critical for interaction with HHR23A UBA(2), synthetic peptides derived from the UBA domains of *S. cerevisiae* (*2*) and an unrelated protein c-cbl (*30*) that is involved in signal transduction in hematopoietic cells were assayed for binding to Vpr in vitro. The amino acid sequence of these UBA domains differs at and around the GFP-loop. The results indicate that the Pro within the GFP-loop of the HHR23A UBA(2) domain is required for interaction with Vpr. Further studies with mutant peptides demonstrate that the Pro within the GFP-loop is necessary but not sufficient for binding to Vpr. Furthermore, the HHR23A UBA(1) domain and the UBA domain from c-cbl do not interact with Vpr, suggesting that the interaction between the HHR23A UBA(2) domain and Vpr is a highly specific one. High-resolution structures of both the wild type and the P333E mutant of UBA(2) were determined by NMR to address the question of how this mutation affects the structure and/or hydrophobic surface of the protein. Comparison of wild-type and P333E mutant structures reveals that substitution of the Pro residue 333 with a Glu results in changes in the shape and charge of the hydrophobic patch predicted to be the Vpr-interacting domain. These changes can account for the loss of the ability to bind Vpr. The structural analysis of a Vpr-interacting protein and identification of a Vpr-interacting interface provide a rationale for the future development of therapeutic reagents that interfere with Vpr function.

MATERIALS AND METHODS

Expression of GST Fusion Proteins in Bacteria. The HIV-1 Vpr GST fusion protein was expressed and purified as described with some modifications (*18*). Briefly, the HIV-1_{NL4-3} coding sequence, inserted into the pGEX-KG expression plasmid, was expressed in *Escherichia coli* BL21 (DE3). Following disruption of the cells by sonication and centrifugation, the cells were dialyzed overnight at 4 °C against PBS. The dialysate was then applied directly to a prepacked glutathione sepharose column (Pharmacia Biotech) and purified according to the manufacturer's instructions. Fractions were collected and analyzed by SDS−polyacrylamide gel electrophoresis (PAGE). Proteins were visualized by staining the polyacrylamide gels with Coomassie Blue or by transferring the proteins to nitrocellulose and staining with Ponceau S. These assays confirmed that all fractions were free of contaminating bacterial proteins. Those fractions containing the bulk of the GSTVpr fusion protein were pooled and dialyzed overnight at 4 °C against PBS. Protein concentration was determined by the Bradford method. Purified proteins were stored at −70 °C.

Synthesis and Preparation of Peptides. Peptides corresponding to UBA domains were biochemically synthesized by Fmoc (fluorenylmethoxycarbonyl)/*tert*-butyl based solid-phase peptide chemistry method on an ABI 431A (Perkin-Elmer) peptide synthesizer utilizing a single coupling program to carry out the chain assembly. The crude peptide was purified to 95% homogeneity by preparative reverse-phase HPLC on a Vydac C18 column. The composition and purity of each peptide was assessed using electro-spray ionization (ESI) mass spectrometry and amino acid analysis. For detection purposes, all peptides were synthesized with a biotin molecule at their N-termini. The sequences of the peptides used in these studies are listed below with mutations indicated in boldface type.

HHR23A UBA(2): QEKEAIERLKALGFPESLVIQAYFACEKNENLAANFLLSQNFDDE; HHR23A UBA(2)P333E: QEKEAIERLKALGF**E**ESLVIQAYFACEKNENLAANFLLSQNFDDE; HHR23A UBA(2)$_{333}$GFP/ERD$_{335}$: QEKEAIERLKALGF**ERD**LVIQAYFACEKNENLAANFLLSQNFDDE; HHR23A UBA(1): SEYETMLTEIMSMGFERERVVAALRASYNNPHRAVEYLLTGIPGS; *S. cerevisiae* UBA(2): EDDQAISRLCELGFERDLVIQVYFACDKNEEAAANILFSDHAD; *S. cerevisiae* UBA(2)E370P: EDDQAISRLCELGF**P**RDLVIQVYFACDKNEEAAANILFSDHAD; *S. cerevisiae* UBA(2)$_{370}$PES$_{372}$: EDDQAISRLCELGF**PES**LVIQVYFACDKNEEAAANILFSDHAD; UBA c-cbl: SPQLSSEIENLMSQGYSYQDIQKALVIAQNNIEMAKNILRE.

Samples for binding assays were prepared by dissolving the purified lyophilized peptide in 10 mM Hepes, pH 6.8 (1 mg/mL). In some cases, the pH was adjusted to 7.0 by the addition of 1 N NaOH. Aliquots were stored at −70 °C and thawed just prior to use.

In Vitro Binding Assay. This assay is a slightly modified version of a binding assay that we have previously developed and tested with all of the relevant controls (*18*). Glutathione Sepharose 4B (Pharmacia Biotech) was prepared by washing 0.5 mL of the suspension obtained from the manufacturer three times in prebinding buffer (100 mM NaCl, 25 mM Tris-

HCl [pH 7.5], 0.1% Triton X-100) (1 mL/wash) and then resuspending it in prebinding buffer in a total volume of 2 mL. GSTVpr or GST (10 μg/reaction) was added to 100 μL of the glutathione Sepharose suspension. After samples were incubated at room temperature (30−60 min), prebound complexes were washed once in prebinding buffer (1 mL/reaction) and then twice with buffer A (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM DTT). Peptides were diluted 4-fold in buffer A, and 50 μg was added to each reaction. After a 30-min incubation at room temperature, samples were washed three times in buffer B (20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 1 mM DTT, 0.5% NP40) (1 mL/wash). Samples were then incubated in blocking buffer [1× TNT (20 mM Tris, 500 mM NaCl, pH 7.5, 0.3% Tween 20 (v/v), 4.0% nonfat dry milk (w/v), 1.0% BSA (w/v)] (0.2 mL/sample) for 30 min at room temperature, and then washed three times in buffer B. After a 30-min incubation in a streptavidin horseradish peroxidase conjugate (Amersham Life Science Inc.) [diluted 1/5000 in TNT that was 0.2% in BSA (w/v)], samples were washed three times in buffer B. 1-Step Turbo TMB-ELISA solution (Pierce Chemical Co.) was then added to each sample (0.25 mL/sample). After a 10-min incubation at room temperature, 150 μL of sulfuric acid (1.2 N) was added to each sample, and the absorbance was read at 450 nm in a spectrophotometer.

Preparation of the UBA(2) Domain. A DNA fragment encoding the UBA(2) domain (residues 319−363) was amplified by PCR from full-length HHR23A gene (*18*) and inserted into the pGEX-2T vector (Pharmacia) between the *Bam*HI and *Eco*RI restriction sites. *E. coli* strain BL21(DE3)-pLysS was then transformed with the plasmid, and the transformed bacteria were cultured in either LB or M9 media. To prepare ^{15}N- or ^{15}N- and ^{13}C-labeled protein, $^{15}NH_4Cl$ or [$^{13}C_6$]glucose was used as sole nitrogen or carbon source in the growth medium. Expression of GST-UBA(2) fusion protein was induced with 1 mM isopropyl-β-D-thiogalactopyranoside when the absorbance of the culture reached approximately 0.7 optical density units at 550 nm. The temperature was lowered from 37 to 30 °C during the expression to increase the yield of soluble fusion protein. The cells were harvested 3 h (4.5 h for the preparation of labeled protein in M9 medium) after induction and resuspended in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.375 M sucrose (15 mL/L of culture). Cells were lysed by adding 1 vol of BugBuster (Novagen), 1 mg/mL lysozyme (Sigma), and 0.5 U of Benzonase (Novagen). The GST fusion protein was cleaved using biotinylated thrombin (Novagen). The protease was then removed with streptavidin agarose, and the UBA(2) protein was finally purified by an anion-exchange chromatography step (high performance Q sepharose, Pharmacia). The final purified UBA(2) protein consisted of residues 319−363 of the HHR23A protein and an N-terminal Gly-Ser peptide derived from the expression vector. Typically, between 5 (^{15}N- and ^{13}C-labeled) and 12.5 mg (^{15}N-labeled) of purified UBA(2) protein was obtained from 1 L of M9 medium. Initial NMR studies for the wild-type protein as well as the NMR analysis of the UBA(2) mutant P333E were performed using peptides derived from chemical synthesis as previously described (*29*).

NMR Analysis and Structure Calculation of the Wild-Type UBA(2) Domain and the Mutant P333E. NMR samples of wild-type UBA(2) contained between 1.5 and 3 mM protein in 50 mM sodium phosphate, pH 6.5, 100 mM sodium chloride, 2 mM dithiothreitol-d_8 (DTT), 0.02% sodium azide, and 5% ^2H_2O. All spectra were acquired at 300 K with a Bruker DRX 500 or DRX 600 spectrometer. A series of three-dimensional experiments (^{15}N-HSQC-NOE-HSQC, ^{15}N-HSQC TOCSY, CBCA(CO)NH, CBCANH, HBHA(CO)NH, HBHANH, HCCH-COSY, and HCCH-TOCSY) were acquired for the assignments of the ^1H, ^{15}N, and ^{13}C resonances (*31*). All NMR data was processed using XWIN-NMR (Bruker) and analyzed with the software XEASY (*32*). Distance information was obtained from ^{15}N- or ^{13}C-edited 3D NOESY experiments as well as a series of 2D NOESY experiments with mixing times ranging from 80 to 300 ms. NOE connectivities obtained from 3D NOESY experiments were categorized into strong, medium, weak, and very weak, corresponding to upper limits for proton−proton distances of 3.0, 4.0, 4.5, and 5.0 Å, respectively. 2D NOESY spectra were analyzed and integrated using the software packages XEASY (*32*) and SPSCAN (*33*). No stereospecific assignments of methylene protons or methyl groups were used. Upper limit distance restraints were then generated from the NOE volumes using DYANA (*34*). In total, 779 NOE-derived distance restraints (173 intraresidue, 196 sequential, 219 medium range, and 191 long range) were used for the wild-type UBA(2). In the final stages of refinement, 17 hydrogen bonds were introduced based on a regular pattern of NOEs of the type $H_{\alpha i}H_{\beta i+3}$ and $H_{\alpha i}NH_{i+3}$, as well as carbon and proton chemical shift information suggesting a regular α-helix.

The proton resonances of the UBA(2)P333E were assigned by analysis of a series of 2D ^1H NMR experiments composed of TOCSY, DQF-COSY, and NOESY. NOE connectivities were derived from a series of 2D NOESY experiments acquired with mixing times ranging from 100 to 350 ms. Distance restraints were generated as described above. A total of 826 NOE-derived distance restraints (204 intraresidue, 182 sequential, 226 medium range, and 214 long range) were used for the structure calculation of UBA(2)P333E. Although almost all NOE-derived distance restraints were derived from the 2D NOESY data acquired on the mutant sample, an additional 148 distance restraints were taken from the wild-type UBA(2) 3D data set due to spectral overlap in the 2D NOESY spectra of the mutant. All of these were located within the hydrophobic core and were far from the site of mutation. Identical proton chemical shifts between the wild-type and the mutant protein indicate that the structural environment for those protons is the same, and use of those restraints does not bias the structure of the mutant protein.

The structures of wild-type UBA(2) and its mutant P333E were calculated from an extended peptide conformation with a simulated annealing protocol using XPLOR (*35*). The simulated annealing protocol consisted of 35 ps dynamics at 2000 K (integration time steps of 5 fs) and 35 ps slow cooling (10 K/cycle). Final minimization was performed using 50 steps of Adopted Raphson Newton algorithm (Quanta98, MSI). For each protein, 100 structures were calculated, of which 18 (wild-type) and 21 (mutant P333E) were selected based on lowest total and NOE energy.

Coordinates for the 18 lowest energy structures of UBA(2) (accession code 1DV0) and UBA(2)P333E (accession code 1F4I) have been deposited in the RCSB protein data bank.

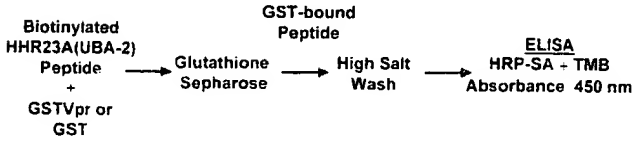


FIGURE 2: Schematic of the in vitro binding assay (*18*). Binding reactions were performed in the following manner. Ten micrograms of GSTVpr or GST bound to glutathione sepharose was mixed with 50 μg of biotinylated UBA(2) peptide and incubated at room temperature for 1 h. GST-containing complexes were selectively recovered and incubated with streptavidin conjugated to horseradish peroxidase. The presence of bound biotinylated peptide was determined by reading the absorbance at 450 nm after the addition of tetramethyl benzidine substrate solution (1-Step Turbo TMB-ELISA, Pierce Chemical Co.) and sulfuric acid.

RESULTS AND DISCUSSION

Binding of Vpr to UBA Domains Is Specific to the UBA-(2) of HHR23A. Our previous studies showed that Vpr binds specifically to UBA(2) of HHR23A and HHR23B. Since the UBA domain is a common structural motif, the binding to Vpr of a variety of other UBA domains was tested using a binding assay based on one that we have previously described (*18*) (Figure 2). Binding of GSTVpr-associated proteins was detected by ELISA. The amino acid sequences and alignment of the peptides that were tested are shown in Figure 3, panel A. The results of the in vitro binding assays using peptides derived from the internal UBA domain of HHR23A [UBA-(1)], *S. cerevisiae* Rad23A UBA(2), and the UBA domain of an unrelated protein, c-cbl (*30*), which is involved in signal transduction in hematopoeitic cells, are shown in Figure 3, panel B. None of these peptides show binding to GSTVpr in this assay. The *S. cerevisiae* Rad23 UBA(2) domain contains several hydrophilic residues in the vicinity of the GFP-loop. Western blot analysis (*29*) also did not show any detectable binding of *S. cerevisiae* Rad23 UBA(2) to GSTVpr (unpublished results). The absence of binding to the UBA domains of c-cbl, *S. cerevisiae* Rad23, as well as the UBA(1) of HHR23A indicates that the interaction between the HHR23A UBA(2) domain and Vpr is a highly specific one and not simply the result of the promiscuous interaction of Vpr with a domain found within many proteins with diverse cellular functions. Furthermore, these results support the prediction that the presence of hydrophilic residues at and around the region of the GFP-loop might preclude Vpr binding and lend support to the proposal that the GFP-loop is the interface for Vpr interaction.

Identification of Critical Residues on HHR23A for Binding to Vpr. As a next step in characterizing the amino acids that are critical for interaction with HHR23A UBA(2), we performed in vitro binding studies with synthetic mutant peptides derived from the UBA(2) domains of *S. cerevisiae* and HHR23A. A comparison of the amino acid sequence of the UBA domains in the vicinity of the GFP-loop indicated the presence of several charged residues in the *S. cerevisiae* UBA(2) domain at and around the GFP-loop that might contribute to the loss of binding to GSTVpr. In particular,

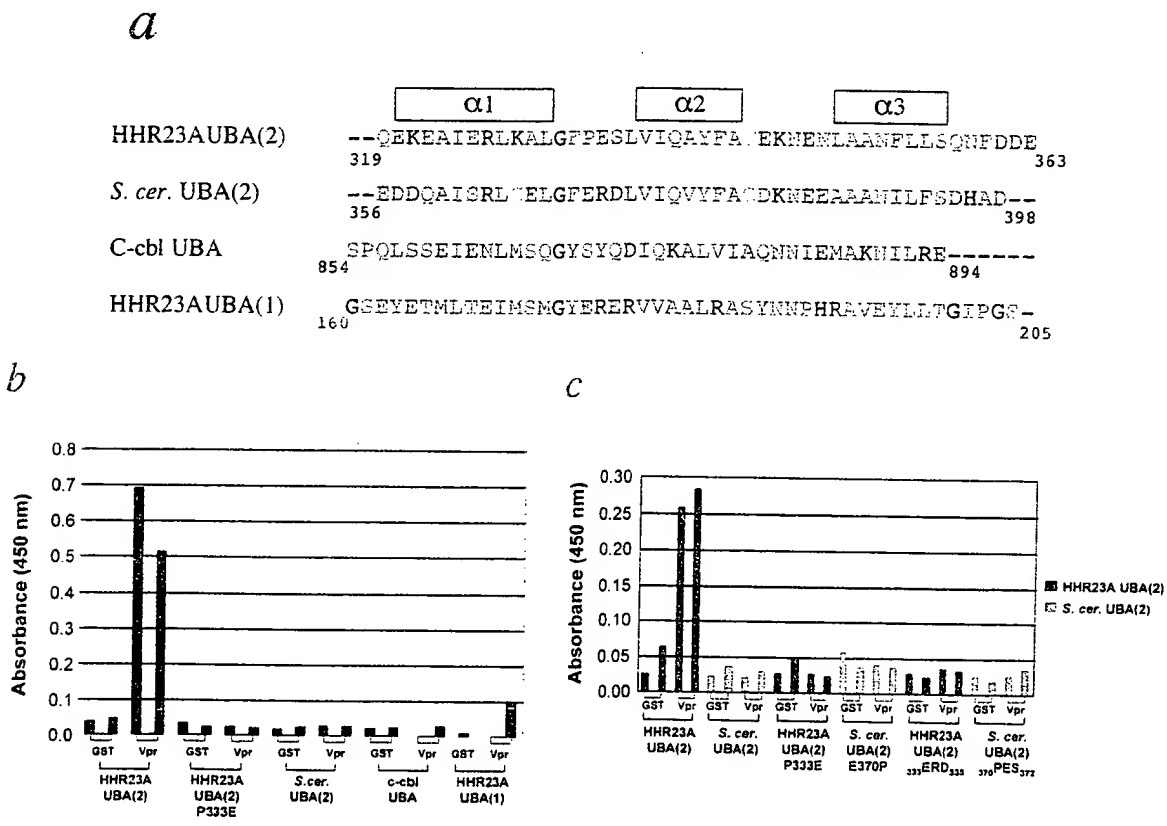


FIGURE 3: (A) Amino acid sequences and alignment of UBA domains: HHR23A UBA(2), *S. cerevisiae* Rad23A UBA(2), c-cbl UBA, and HHR23A UBA(1). The amino acids that comprise α-helices 1, 2, and 3 are indicated above the sequence. The sequence of amino acids are colored green (hydrophobic), red (negatively charged), orange (polar uncharged), and blue (positively charged). (B) Results of binding assays with HHR23A UBA(2), HHR23A UBA(2) P333E, *S. cerevisiae* Rad23A UBA(2), c-cbl UBA, and HHR23A UBA(1). (C) Results of binding assays with HHR23A and *S. cerevisiae* Rad23 UBA(2) mutant peptides. Binding reactions were performed as described in the Materials and Methods section. Data shown are values obtained from duplicate samples from a typical experiment.

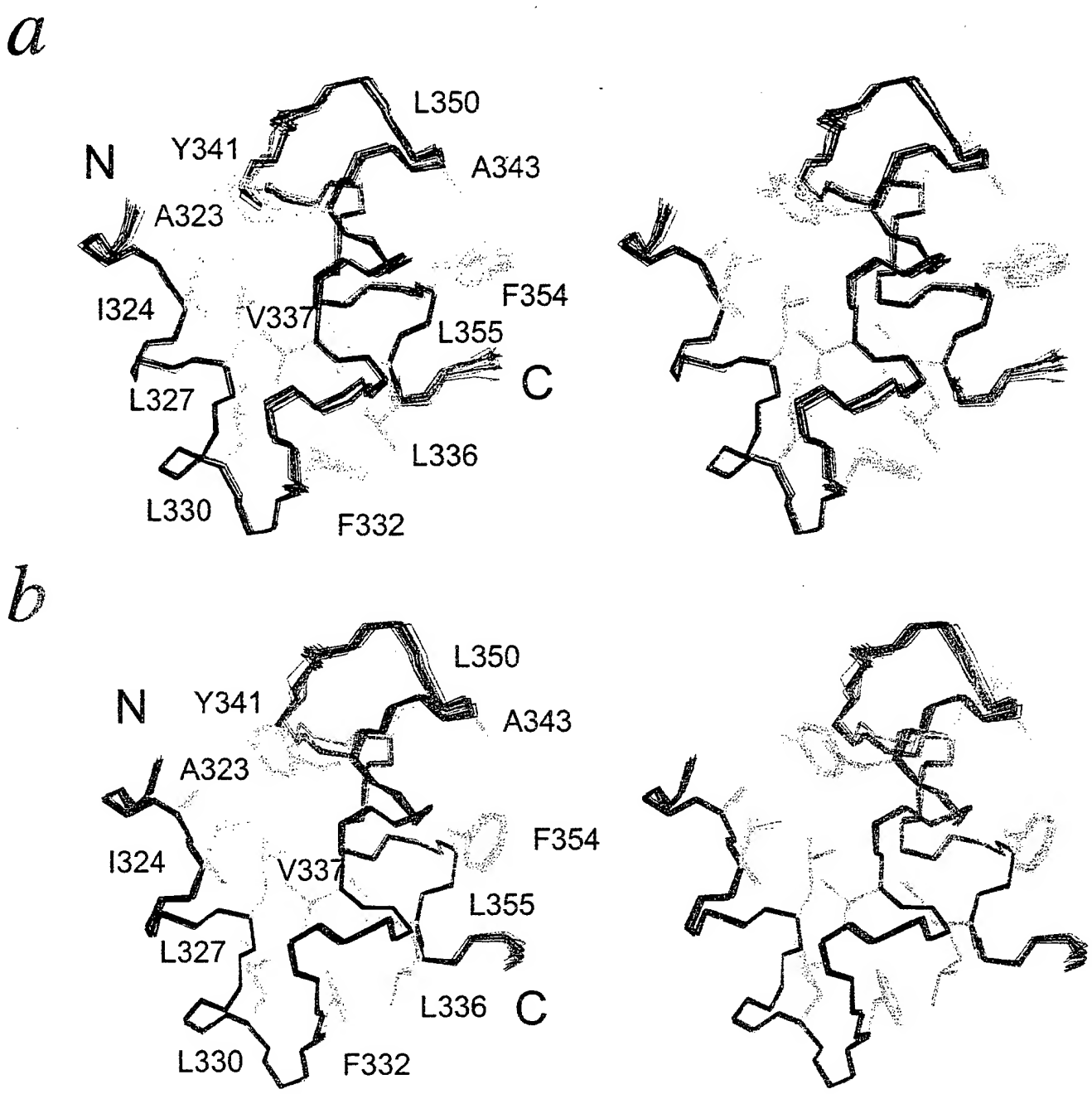


FIGURE 4: Stereoviews of superpositions of the lowest energy structures of (A) wild-type UBA(2) (18 structures) and (B) UBA(2)P333E (21 structures). Backbone atoms for residues 320−357 are shown. The C-terminal residues 358−363 are disordered and are not shown here. Side chain atoms for the residues contributing to the hydrophobic core are displayed in green.

we predicted that the Pro residue at position 333 of HHR23A UBA(2) might be critical for interaction with Vpr. To assess the contribution of this residue to the HHR23A UBA(2)/Vpr interaction, a mutant HHR23A UBA(2) peptide containing a glutamic acid residue at position 333 [HHR23A UBA(2)P333E)] was chemically synthesized and assayed for binding. This P333E substitution completely abolished binding of the peptide to GSTVpr (Figure 3, panel C). The HHR23A UBA(1) domain contains a glutamic acid residue at the position corresponding to position 333 in the HHR23A UBA(2) domain, and the c-cbl UBA domain contains a serine at that position. These results indicate that the proline at position 333 of HHR23A UBA(2) is necessary for interaction with GSTVpr.

Further studies with mutant synthetic UBA(2) peptides demonstrate that the Pro at position 333 is critical for binding but not sufficient on its own for the interaction of UBA(2) with GSTVpr. Neither an E370P substitution nor a $_{370}$PES$_{372}$ substitution made within the *S. cerevisiae* UBA(2) confers upon those peptides the ability to bind GSTVpr (Figure 3, panel C). These in vitro binding results demonstrate the importance of the hydrophobic GFP-loop region for interaction of UBA(2) with Vpr.

Refined High-Resolution Structures of Wild-Type UBA-(2) of HHR23A. To further explore the role of the critical P333 residue on binding of UBA(2) to Vpr, we undertook to determine the structure of UBA(2) P333E. During the assignment of this mutant protein using 2D proton NMR data recorded from a synthetic peptide, we discovered a mismatch between the assignments for the wild-type UBA(2) (*29*) and the mutant. To resolve the conflict, the structure of HHR23A UBA(2) was redetermined using heteronuclear NMR spec-

troscopy on a ^{13}C-,^{15}N-labeled protein. The high resolution structure was necessary for the structure comparison, to analyze the small structural changes that might be caused by the point mutation of Pro333 to Glu. For the preparation of labeled protein, the cDNA of UBA(2) encoding the residues 319−363 was cloned into the expression vector pGEX-2T. Analysis of the isotope-edited 3D data revealed a misassignment in the original data due to spectral overlap. Specifically, NOEs were incorrectly assigned between the side chain of Leu330 and the aromatic ring protons of Tyr341. The correct NOE connectivities are found to be between the side chains of Ile324 and Tyr341 as well as Glu348 and Tyr341. For the highly refined structure of the wild-type UBA(2) domain 779 NOE-derived distance restraints (vs 380 in the previously reported structure) were obtained. Simulated annealing calculations were employed to generate an ensemble of 18 structures (Figure 4, panel A) consistent with the NMR data. The structures exhibit good covalent geometry and have no NOE violations greater than 0.5 Å. Figure 4, panel A, shows the quality of the calculated structures with a root-mean-square (rms) deviation between the atomic coordinates of the backbone and all heavy atoms with respect to the average coordinates for the residues within the three α-helices (residues 320−330, 337−344 and 350−357) of only 0.178 ± 0.037 Å and 0.686 ± 0.064 Å, respectively. Complete restraint and structural statistics are presented in Table 1.

In the refined structure, the three helix bundle is more compact and the orientation of the interhelical angles changed slightly (Figure 4, panel A). Interhelical angles were determined using the program RIBBONS2.6 (*36*). Helices 1 and 2 are arranged at an angle of 120° (equivalent to −60°) and helices 2 and 3 exhibit an angle of 126° (or −54°). These values are close to the optimum of 52° reported for a "knob in hole" packing of the side chains. In contrast, the interhelical angle between helices 1 and 3 is only about 30°. A tight "knob in hole" packing for the side chains of these two helices is not required due to the larger distance between them. The resulting change in the structure of the UBA(2) domain is mainly a shift of the first α-helix of the three-helical bundle along its axis of about 3 Å. The orientation of the first helix with respect to the other two helices is only slightly affected (about 15°).

The structure of the UBA(2) domain is well-ordered from residues Glu320 to Ser357. In contrast to the low resolution structure of UBA(2) (*29*), a full set of NOEs of the type $H_{\alpha i}H_{\beta i+3}$ and $H_{\alpha i}NH_{i+3}$, were found for residues Glu320 to Leu330, showing that the first helix is extended to the N-terminal end of the UBA(2) domain. The C-terminal residues Asn359 to Asp363 show no medium or long range NOE connectivities and are disordered. Carbon and proton chemical shifts for these residues are similar to those found for random coil conformations. The first loop connecting helices 1 and 2 is well-ordered, and the residues Gly331 to Leu336 show many NOE connectivities to residues located in the hydrophobic core. NOE connectivities for P333 clearly indicate that the Pro residue adopts a trans conformation. The second loop, consisting of the residues Glu345 to Asn349, is far more exposed on the surface. Line broadening for several residues in this loop, namely Lys346 and Glu348, indicates that this loop is dynamic, likely due to slow chemical exchange processes.

Table 1: Input Data for Structure Calculation and Structural Statistics UBA(2) [18 Structures] and UBA(2)P333E [21 Structures]

Distance Restraints	UBA(2)	UBA(2)P333E
total	796	843
NOE-derived		
intraresidue	173	204
sequential ($\|i-j\|=1$)	196	182
medium-range ($\|i-j\|\leq 4$)	219	226
long-range ($\|i-j\|>4$)	191	214
hydrogen bonds		
medium-range (within-helices)	17	17
Structure Statistics		
RMSD from experimental distance restraints (Å)	0.047 ± 0.001	0.052 ± 0.0003
RMSD from idealized geometry		
bonds (Å)	0.005 ± 0.00004	0.005 ± 0.00005
angles (deg)	0.616 ± 0.007	0.617 ± 0.004
energies (kcal mol^{-1})		
E_{NOE}	11.7 ± 0.4	9.8 ± 0.5
E_{bond}	17.1 ± 0.3	18.3 ± 0.4
E_{angles}	74.9 ± 1.6	74.8 ± 1.0
E_{L-J}[a]	−262± 9	−270 ± 6
RMSD from avg structure (Å)		
backbone atoms in secondary structure[b]	0.18 ± 0.04	0.12 ± 0.03
backbone atoms for residues 3−9	0.21 ± 0.06	0.16 ± 0.07
all heavy atoms in secondary structure[b]	0.69 ± 0.06	0.66 ± 0.08
all heavy atoms for residues 3−39	0.72 ± 0.07	0.71 ± 0.07
Procheck analysis		
residues in most favored region (%)	79.7	76.1
residues in additional allowed region (%)	15.9	20.9
residues in generously allowed regions (%)	3.7	2.6
residues in disallowed regions (%)	0.8	0.4

[a] Lennart−Jones energy was calculated after the accepted structures were subjected to a final minimization (50 steps Adopted Raphson Newton algorithm, QUANTA98) using the CHARMM force field and the NOE-derived distance restraints. [b] Secondary structure is composed of residues 3−12, 19−25, and 32−39. These residues are numbered 321−330, 337−343, and 350−357 in HHR23A.

One interesting feature of the UBA(2) domain is the presence of large hydrophobic patches at the protein surface (Figure 5). Several hydrophobic residues that show contacts with the hydrophobic core are partially exposed on the surface. The largest area is made up of residues Leu327, Ala329, and Leu330 in helix 1, residues Gly331, Phe332, Pro333, and Leu336 in the loop 1, and residues Ala351 and Leu355 in helix 3, covering a total of about 596 Å^2 (epitope 1). This corresponds to approximately 18% of the total solvent accessible surface area of 3288 Å^2. A second hydrophobic region (epitope 2) located on the opposite side of the UBA(2) molecule consists of the residues Ala323, Ile324 (helix 1), Ile338, Phe341, Ala342 (helix2), Leu349 (helix3), and Phe359 (C-terminus). The total surface area is 555 Å^2, but in contrast to the hydrophobic epitope 1, this area is not as compact and the residues do not form a consecutive epitope. Such large hydrophobic surface patches are rather unusual and as previously noted are likely binding sites for other proteins. The site of interaction of HIV-1 Vpr with the UBA(2) domain was therefore suggested to occur

a

319QEKE..ERLK..G..ESLV..QAY..ACEKNENLA..L SQN..DDE363

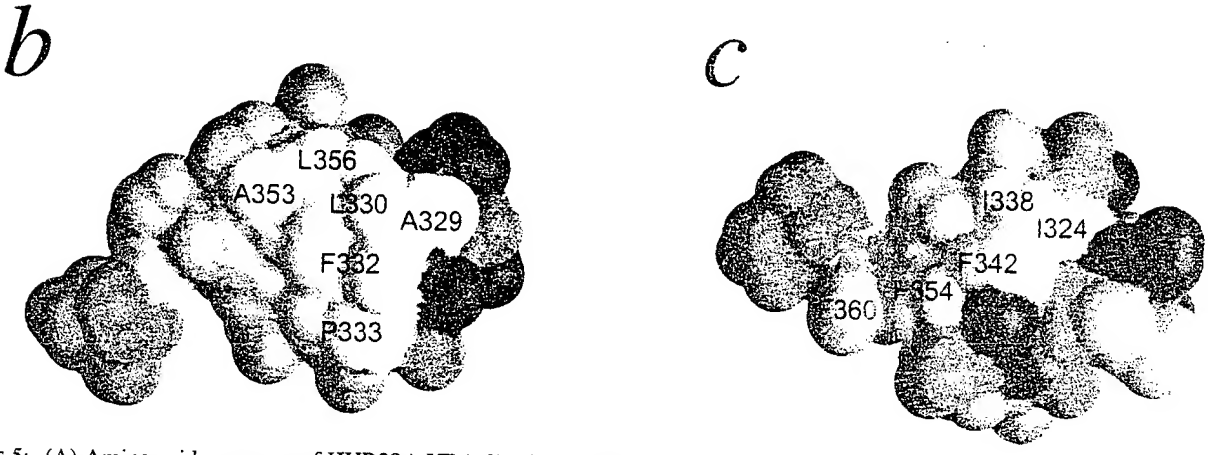


FIGURE 5: (A) Amino acid sequence of HHR23A UBA(2) with the amino acids that comprise the hydrophobic epitope 1 and 2 highlighted in yellow and blue, respectively. (B) Hydrophobic surface view of UBA(2) showing the hydrophobic epitope 1, which is the proposed binding site for HIV-1 Vpr. (C) View of the surface representation as in panel B but rotated upon the *x*-axis about 180°, showing the hydrophobic epitope 2. Hydrophobic amino acids (A, G, F, I, L, M, P, V) are colored in white, negatively charged amino acids (D, E) in red, positively charged in blue (K, R), and polar residues (N,Q, S, T) are in orange.

between the hydrophobic region of Vpr and epitope 1. Epitope 2 is a potential site for other protein−protein interactions in the XPC complex.

Comparison of the Structures of Wild-Type UBA(2) and Mutant P333E. As discussed, a single point mutation of the residue Pro333 to a Glu has been shown to completely abolish the binding of HIV-1 Vpr to UBA(2). To investigate whether the impaired binding is due to a change in structure or the change in a critical interaction with the amino acid side chain, we determined the structure of the UBA(2) mutant P333E (Figure 4, panel B). A synthetic peptide was used for the NMR analysis. For the final structure calculations, 826 NOE-derived distance restraints were used, of which 204 are intraresidual, 182 are sequential, 226 are medium range, and 214 are long-range NOE connectivities (Table 1). Twenty-one of 100 calculated structures were selected based on lowest total and NOE energy. Figure 4, panel B, shows the high structural precision of 0.121 ± 0.031 Å for the backbone atoms and 0.661 ± 0.080 Å for the heavy atoms for the three helices. The high-resolution structures for the wild-type and mutant UBA(2) domain were of similar resolution enabling us to do a detailed structure/function analysis.

The coordinates for the backbone atoms within the three helices of the wild-type UBA(2) and mutant P333E are nearly identical, showing that the only structural changes between the wild-type and mutant UBA(2) occur locally around the mutated residue 333 (Figure 4). Comparison of the proton chemical shifts of the wild-type and the mutant resonances also indicates that there is no change in the overall structure (data not shown). Differences in the proton chemical shift larger than 0.1 ppm are observed for the residues close to the site of mutation, i.e., for the residues Gly331, Phe332, Glu334, and Ser335.

Figure 6 shows the superposition of the structural ensemble of wild-type UBA(2) and the mutant P333E for the residues Leu327 to Val337 around the GFP-loop. The structures were superimposed using the backbone atoms within the three helices. Structural changes within loop 1 between wild-type and mutant are small but significant. The root-mean-square deviation to the mean within loop 1 is about 0.2 Å for wild type and P333E. However, the difference in the positions of the amide nitrogens of residues 333 and 334 are about 1.2 and 0.8 Å respectively. This clearly indicates that the structural changes observed are significant within the accuracy of the NMR data. The largest change in main chain conformation occurs at the site of mutation. The backbone torsion angles for the residue Phe332 preceding the mutated residue are different for the wild-type and the mutant protein. The difference of the φ angle for Phe332 angle is small, about −152° for the wild-type and −168° for the mutant. In contrast, the ψ angle for Phe332 is drastically altered from about −138° for the wild type to −77° for the mutant P333E. This results in a different main chain conformation for the loop 1 for wild type and mutant from residue Phe332 to residue Ser335. Starting with residue Val337 the main chain and side chain coordinates can again be superimposed within the error limits of the NMR data. The differences in main chain torsion angle also lead to different packing for the residues in contact with residue 333. Pro333 of the wild-type UBA(2) shows contacts to other residues participating in the hydrophobic core, namely, residues Phe332 and Leu336. The χ_1 torsion angle of residue Phe332 is changed from −60° to −52°, and the χ_2 torsion angle is altered from

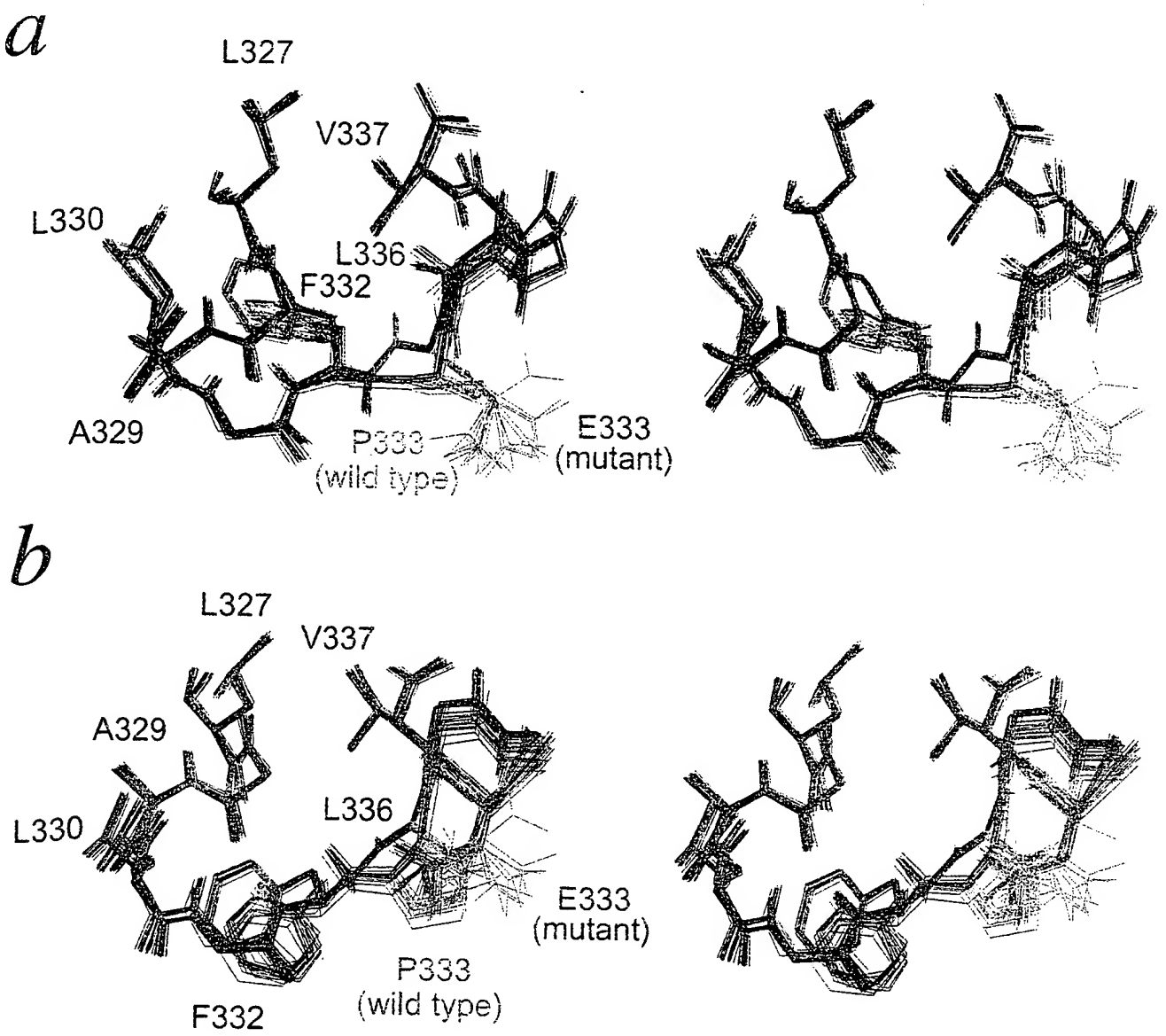


FIGURE 6: (A) Comparison of the structure of loop 1 (residues Leu327 to Val337) of UBA(2) (black) and UBA(2)P333E (blue). (B) Same as in panel A but rotated around the *x*-axis by approximately 90°. The site of mutation is illustrated by different colors for the respective side chains, red for the wild-type Pro333 and green for the mutant Glu333. The superposition of the structures reveals that the differences between the loop conformations of the wild-type and mutant proteins are larger than the rms deviation within their structure ensembles. Residues farther away from the mutation, e.g., Leu327 and Val337, have identical side chain conformations in the wild-type and mutant protein domain structures. The major structural change occurs at the backbone ψ torsion angle for residue Phe332, which leads to a different side chain conformation of the Phe residue and a different orientation of the mutated residue 333.

107° to 152° for the wild-type vs P333E mutant proteins, respectively. Differences of a similar order of magnitude are observed for residue Leu336 as well. The torsion angles χ_1 and χ_2 are changed by about 16° (from 68° for wild-type to 84° for P333E) and 54° (from −172° for the wild-type to −118° for P333E), respectively. Residue Leu330, which has no direct contact to residue 333, exhibits only small changes for the side chain orientation as an effect of the packing against residue Phe332. The side chain torsion angles for residues Leu327 and Val337, both of which are packed in the hydrophobic core, are identical within the error limits of the NMR data.

Comparison of Structural Data with the Results of S. cer. Rad23A Mutagenesis. A NOESY spectrum of the *S. cerevisiae* Rad23A UBA(2) domain showed a very similar NOE pattern to the wild-type UBA(2) domain (data not shown). This indicates that, like the P333E mutant, it exhibits the same overall fold and secondary structure as the HHR23A UBA(2). The *S. cerevisiae* Rad23A UBA(2) domain does not bind to Vpr, as discussed. In comparison to the UBA(2) domain of HHR23A, the UBA(2) domain of *S. cerevisiae* differs in the amino acid sequence in the vicinity of the GFP-loop, specifically Ala329 is substituted by a negatively charged Glu, Pro333 is exchanged by a Glu, the subsequent amino acid Glu334 is charge-reversed with an Arg residue, and the Ser335 residue is replaced by a negatively charged Asp. Analysis of a UBA(2) chimera derived from HHR23A and *S. cerevisiae* demonstrated that the Pro residue within the GFP-loop is insufficient for binding to Vpr (Figure 3, panel C). Replacement of the Glu residue found at position 370 in the UBA domain of *S. cerevisiae* [equivalent to position 333 in HHR23A UBA(2)] with a Pro does not restore binding to HIV-1 Vpr, indicating that there are additional determinants for the binding of Vpr. Using the sequence alignment of Figure 3, panel A, we can predict the equivalent hydrophobic epitope 1 for the UBA domain

of *S. cerevisiae*. This analysis shows that in addition to the change to Glu at position 333, two other residues that are part of the hydrophobic patch are different between the two UBA domains, i.e., Ala329 and Leu356, which are replaced by Glu and Phe, respectively. This change of two small hydrophobic residues for a charged and much larger amino acid, respectively, is likely to have a significant effect on the structure of the proposed interaction surface between UBA(2) and Vpr.

SUMMARY AND CONCLUSIONS

The UBA domains are found in a number of proteins besides HHR23A with diverse functions involved in ubiquitination, DNA repair, and signaling pathways via protein kinases (*16*). One of the unusual features of the UBA(2) domain of HHR23A is the presence of large hydrophobic patches at the protein surface. The existence of these large hydrophobic patches strongly suggests that these may serve as binding sites for other proteins. For HIV-1 Vpr, we confirmed by site-directed mutagenesis that one of the hydrophobic patches encompassing Pro333 is critical for the interaction between Vpr and UBA(2) (*18*). To understand the structural basis of the loss in binding upon mutation of Pro 333 to Glu, we determined the structures of the wild-type UBA(2) domain and the mutant P333E to high resolution. A comparison shows that while the overall fold of the UBA domain is not affected by the substitution, there are significant changes that occur locally around the site of mutation. The high accuracy of the structures revealed that substitution of a Pro residue 333 with a Glu resulted in changes of the shape and charge of the hydrophobic patch predicted to be the Vpr-interacting domain. These changes account for the loss of the ability to bind Vpr.

The structural and genetic information developed in the study should facilitate the identification and characterization of cellular proteins that bind through UBA. One hypothesis that can be addressed with the structural knowledge obtained here is that Vpr may compete or alter the interaction between UBA(2) and a normal cellular protein involved in DNA repair signaling processes. To date, despite their ubiquitous occurrence, only two cellular binding partners for UBA domains have been identified. Ubiquitin has been shown to bind to the UBA domain of p62, a ligand of the SH2 domain of $p56^{lck}$ (*37*). However, the very low sequence homology between the UBA domain of p62 and UBA(2) of HHR23A does not suggest a similar interaction between UBA(2) and ubiquitin. Significantly for this work, HHR23A and B have been shown to interact with the 3-methyladenine DNA glycosylase (MPG) protein and have been proposed to serve as universal DNA damage recognition accessory proteins. Furthermore, MPG binds specifically to the UBA(2) of HHR23A. It is therefore possible that HIV-1 Vpr may compete with MPG for binding to HHR23A, thus affecting DNA nucleotide excision repair. This provides support for our hypothesis that the interaction of HIV-1 Vpr with this domain may be important to its functional role in mediating cell cycle arrest.

ACKNOWLEDGMENT

We thank Dr. Thorsten Dieckmann for advice in the early stages of this work. We thank Dr. Mark A. Jarosinski and the Peptide Synthesis group at AmGen (Boulder, Co) for synthesis of the peptides used in the binding assays and the UBA(2)P333E peptide used in the structural studies and Mr. Evan Feinstein for help in manuscript preparation.

REFERENCES

1. Lombaerts, M., Goeloe, J. I., den Dulk, R., Brandsma, J. A., and Brouwer, J. (2000) *Biochem. Biophys. Res. Commun. 268*, 210−215.
2. Watkins, J. F., Sung, P., Prakash, L., and Prakash, S. (1993) *Mol. Cell. Biol. 13*, 7757−7765.
3. Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., Vanderspek, P. J., Bootsma, D., Hoeijmakers, J. H. J., and Hanaoka, F. (1994) *EMBO J. 13*, 1831−1843.
4. Madura, K., and Prakash, S. (1990) *Nucleic Acids Res. 18*, 4737−4742.
5. Guzder, S. N., Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1995) *J. Biol. Chem. 270*, 12973−12976.
6. Sugasawa, K., Masutani, C., Uchida, A., Maekawa, T., Vanderspek, P. J., Bootsma, D., Hoeijmakers, J. H. J., and Hanaoka, F. (1996) *Mol. Cell. Biol. 16*, 4852−4861.
7. Masutani, C., Araki, M., Sugasawa, K., van der Spek, P. J., Yamada, A., Uchida, A., Maekawa, T., Bootsma, D., Hoeijmakers, J. H., and Hanaoka, F. (1997) *Mol. Cell Biol. 17*, 6915−6923.
8. Li, L., Lu, X. Y., Peterson, C., and Legerski, R. (1997) *Mutat. Res. 383*, 197−203.
9. Wang, Z. G., Wei, S. G., Reed, S. H., Wu, X. H., Svejstrup, J. Q., Feaver, W. J., Kornberg, R. D., and Friedberg, E. C. (1997) *Mol. Cell. Biol. 17*, 635−643.
10. Jansen, L. E. T., Verhage, R. A., and Brouwer, J. A. (1998) *J. Biol. Chem. 273*, 33111−33114.
11. Guzder, S. N., Sung, P., Prakash, L., and Prakash, S. (1998) *J. Biol. Chem. 273*, 31541−31546.
12. Mueller, J. P., and Smerdon, M. J. (1996) *Mol. Cell. Biol. 16*, 2361−2368.
13. Schauber, C., Chen, L., Tongaonkar, P., Vega, I., Lambertson, D., Potts, W., and Madura, K. (1998) *Nature 391*, 715−718.
14. Lambertson, D., Chen, L., and Madura, K. (1999) *Genetics 153*, 69−79.
15. Hiyama, H., Yokoi, M., Masutani, C., Sugasawa, K., Maekawa, T., Tanaka, K., Hoeijmakers, J. H. J., and Hanaoka, F. (1999) *J. Biol. Chem. 274*, 28019−28025.
16. Hofmann, K., and Bucher, P. (1996) *Trends Biochem. Sci. 21*, 172−173.
17. Miao, F., Bouziane, M., Dammann, R., Masutani, C., Hanaoka, F., Pfeifer, G. P., and O'Connor, T. R. (2000) *J. Biol. Chem. 275*, 28433−28438.
18. Withers-Ward, E. S., Jowett, J. B. M., Stewart, S. A., Xie, Y. M., Garfinkel, A., Shibagaki, Y., Chow, S. A., Shah, N., Hanaoka, F., Sawitz, D. G., Armstrong, R. W., Souza, L. M., and Chen, I. S. Y. (1997) *J. Virol. 71*, 9732−9742.
19. Emerman, M. (1996) *Curr. Biol. 6*, 1096−1103.
20. Jowett, J. B. M., Planelles, V., Poon, B., Shah, N. P., Chen, M. L., and Chen, I. S. Y. (1995) *J. Virol. 69*, 6304−6313.
21. He, J. L., Choe, S., Walker, R., Dimarzio, P., Morgan, D. O., and Landau, N. R. (1995) *J. Virol. 69*, 6705−6711.
22. Re, F., Braaten, D., Franke, E. K., and Luban, J. (1995) *J. Virol. 69*, 6859−6864.
23. Rogel, M. E., Wu, L. I., and Emerman, M. (1995) *J. Virol. 69*, 882−888.
24. Stewart, S. A., Poon, B., Jowett, J. B. M., and Chen, I. S. Y. (1997) *J. Virol. 71*, 5579−5592.
25. Mahalingam, S., Ayyavoo, V., Patel, M., KieberEmmons, T., and Weiner, D. B. (1997) *J. Virol. 71*, 6339−6347.
26. Di Marzio, P., Choe, S., Ebright, M., Knoblauch, R., and Landau, N. R. (1995) *J. Virol. 69*, 7909−7916.
27. Goh, W. C., Rogel, M. E., Kinsey, C. M., Michael, S. F., Fultz, P. N., Nowak, M. A., Hahn, B. H., and Emerman, M. (1998) *Nat. Med. 4*, 65−71.

28. Poon, B., Grovit-Ferbas, K., Stewart, S. A., and Chen, I. S. Y. (1998) *Science 281*, 266−269.
29. Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S. Y., and Feigon, J. (1998) *Nat. Struct. Biol. 5*, 1042−1047.
30. Blake, T. J., Shapiro, M., Morse, H. C., and Langdon, W. Y. (1991) *Oncogene 6*, 653−657.
31. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, Inc., San Diego.
32. Bartels, C., Xia, T. H., Billeter, M., Guntert, P., and Wuthrich, K. (1995) *J. Biomol. NMR 6*, 1−10.
33. Glaser, R. W., and Wüthrich, K. (1997) http://www.mol.biol.ethz.ch/wuthrich/software/spscan/.
34. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
35. Brünger, A. T. (1992) *X-PLOR* (Version 3.1) Manual, Yale University Press, New Haven and London.
36. Carson, M. (1991) *J. App. Crystal. 24*, 958−961.
37. Vadlamudi, R. K., Joung, I., Strominger, J. L., and Shin, J. (1996) *J. Biol. Chem. 271*, 20235−20237.

BI0017071

ed annealing was done with additional restraints that define the U305 and U300 base triples (Table 2). During all calculations the P4-P6 domain and the docked P1 helix were held rigid. In the resulting model, there are no van der Waals or chemical geometry violations.

Note added in proof: While this work was in press, Golden et al.[32] published a model based upon 5 Å crystallographic data of a Tetrahymena intron lacking the P1 helix and several other peripheral elements known to stabilize the catalytic core. Within the J8/7 region there are substantial discrepancies between the 5 Å model and the results presented here as well as other biochemical data[9,10]. In order for J8/7 to form the appropriate base triples with P1 and P3, it must rotate 180° about the phosphate backbone at A301 and A302 and P3 must translate by more than 20 Å near U300. This suggests that the J8/7 region of the crystallized intron lacking the P1 helix may not be preordered for substrate binding.

Acknowledgments

We thank R. Sousa for the clone of the Y639F mutant T7 RNA polymerase, R.R. Gutell for analysis of group I sequences, and L. Weinstein for helpful comments on the manuscript. S.P.R. was supported by an NIH training grant. This work was funded by an NIH grant, a Beckman Young Investigator Award, a Searle Scholar award, and a Junior Faculty Research Award from the American Cancer Society to S.A.S.

Alexander A. Szewczak, Lori Ortoleva-Donnelly, Sean P. Ryder, Eileen Moncoeur and Scott A. Strobel

Department of Molecular Biophysics and Biochemistry, 260 Whitney Avenue, Yale University, New Haven, Connecticut 06520, USA.

Correspondence should be addressed to S.A.S. *email: strobel@csb.yale.edu*

Received 22 June, 1998; accepted 14 September, 1998.

1. Cech, T.R. & Herschlag, D. In *Catalytic RNA*, vol. 10 (eds Eckstein, F. & Lilley, D.M.J.) 1–17 (Springer, New York; 1996).
2. Cate, J.H. *et al. Science* **273**, 1678–1685 (1996).
3. Waring, R.B., Towner, P., Minter, S.J. & Davies, R.W. *Nature* **321**, 133–139 (1986).
4. Bevilacqua, P.C. & Turner, D.H. *Biochemistry* **30**, 10632–10640 (1991).
5. Pyle, A.M. & Cech, T.R. *Nature* **350**, 628–631 (1991).
6. Strobel, S.A. & Cech, T.R. *Biochemistry* **32**, 13593–13604 (1993).
7. Strobel, S.A. & Cech, T.R. *Science* **267**, 675–679 (1995).
8. Strobel, S.A., Ortoleva-Donnelly, L., Ryder, S.P., Cate, J.H. & Moncoeur, E. *Nature Struct. Biol.* **5**, 60–66 (1998).
9. Ortoleva-Donnelly, L., Szewczak, A.A., Gutell, R.R. & Strobel, S.A. *RNA* **4**, 498–519 (1998).
10. Pyle, A.M., Murphy, F.L. & Cech, T.R. *Nature* **358**, 123–128 (1992).
11. Damberger, S.H. & Gutell, R.R. *Nucleic Acids Res.* **22**, 3508–3510 (1994).
12. Tanner, M.A., Anderson, E.M., Gutell, R.R. & Cech, T.R. *RNA* **3**, 1037–1051 (1997).
13. Michel, F. & Westhof, E. *J. Mol. Biol.* **216**, 585–610 (1990).
14. Wang, J.F. & Cech, T.R. *Science* **256**, 526–529 (1992).
15. Waring, R.B., Davies, R.W., Brown, T.A. & Scazzocchio, C. *Gene* **28**, 277–291 (1984).
16. Narlikar, G.J., Khosla, M., Usman, N. & Herschlag, D. *Biochemistry* **36**, 2465–2477 (1997).
17. Michel, F. & Westhof, E. *Nature Struct. Biol.* **1**, 5–7 (1994).
18. Conrad, F., Hanne, A., Gaur, R.K. & Krupp, G. *Nucleic Acids Res.* **23**, 1845–1853 (1995).
19. Strobel, S.A. & Shetty, K. *Proc. Natl. Acad. Sci. USA* **94**, 2903–2908 (1997).
20. Strobel, S.A., Ortoleva-Donnelly, L., Ryder, S.P., Cate, J.H. & Moncoeur, E. *Nature Struct. Biol.* **5**, 60–66 (1998).
21. Moore, M.J. & Sharp, P.A. *Science* **256**, 992–997 (1992).
22. Ortoleva-Donnelly, L., Kronman, M. & Strobel, S.A. *Biochemistry* **37**, 12933-12942 (1998).
23. Moser, H.E. & Dervan, P.B. *Science* **238**, 645–650 (1987).
24. Beal, P.A. & Dervan, P.B. *Science* **251**, 1360–1363 (1991).
25. Lehnert, V., Jaeger, L., Michel, F. & Westhof, E. *Chem. Biol.* **3**, 993–1009 (1996).
26. Pyle, A.M. *Science* **261**, 709–714 (1993).
27. Christian, E.L. & Yarus, M. *J. Mol. Biol.* **228**, 743–758 (1992).
28. Christian, E.L. & Yarus, M. *Biochemistry* **32**, 4475–4480 (1993).
29. Piccirilli, J.A., Vyle, J.S., Caruthers, M.H. & Cech, T.R. *Nature* **361**, 85–88 (1993).
30. Weinstein, L.B., Jones, B.C.N.M., Cosstick, R. & Cech, T.R. *Nature* **388**, 805–808 (1997).
31. Beren, C., Steicher, B., Schroeder, R. & Hillen, W. *Chem. Biol.* **5**, 163–175 (1998).
32. Golden, B. L.,Gooding, A. R., Podell, A. R. & Cech, T. R. *Science* **282**, 259–264 (1998).

Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr

The HIV-1 protein Vpr is critical for a number of viral functions including a unique ability to arrest T-cells at a G2/M checkpoint and induce subsequent apoptosis. It has been shown to interact specifically with the second UBA (ubiquitin associated) domain found in the DNA repair protein HHR23A, a highly evolutionarily conserved protein. This domain is a commonly occurring sequence motif in some members of the ubiquitination pathway, UV excision repair proteins, and certain protein kinases. The three dimensional structure of the UBA domain, determined by NMR spectroscopy, is presented. The protein domain forms a compact three-helix bundle. One side of the protein has a hydrophobic surface that is the most likely Vpr target site.

Vpr mediates a number of different functions during the HIV life cycle[1], including nuclear import[2] and the induction of cell cycle arrest in the G2 phase[3–6]. While the exact mechanism of Vpr in induction of cell cycle arrest has not yet been elucidated, it has been shown that Vpr is both necessary and sufficient to cause accumulation of cells in G2 phase[3]. The arrested HIV-1 infected T-cells subsequently undergo cell death by apoptosis. It has been hypothesized that this can have a significant effect in HIV pathogenesis by preventing efficient activation and subsequent clonal expansion of CD4+ T-cells[7,8]. The Vpr mediated cell cycle arrest is in some aspects similar to the arrest of cells at certain cell cycle checkpoints as a result of exposure to DNA damaging agents[9]. Using a yeast two-hybrid assay, it has recently been shown that Vpr also interacts with the DNA repair protein HHR23A[10], the human homolog of Rad23. HHR23A is an acidic 40,000 M_r protein that shares extensive homology with the protein encoded by the *Saccharomyces cerevisae RAD23* gene[11]. The Rad23 protein is involved in nucleotide excision repair and is up-regulated if the cell is exposed to DNA-damaging agents[11]. All of the Rad23 homologs identified to date share a common domain structure with a so-called ubiquitin-like region at the N-terminus and two highly conserved regions of ~50 amino acids, termed the ubiquitin associated (UBA) domains[12], in the center and at the C-terminus (Fig. 1*a*). The 45 amino acid C-terminal UBA domain (UBA(2)) was identified as the Vpr-interacting domain of HHR23A by an *in vitro* binding assay[10]. UBA domains are a very common sequence motif in proteins of the ubiquitination pathway, UV excision repair proteins and certain protein kinases[12]. Although the specific role of this protein domain is so far not known, it has been suggested that UBA domains are involved in conferring target specificity to multiple enzymes of the ubiquitination system. In order to further our understanding of the interaction of the UBA domain with Vpr and other proteins, we have determined the structure of UBA(2) of HHR23A using homonuclear NMR spectroscopy. This is the first three-dimensional structure of a UBA domain.

Withers-Ward *et al.*[10] have shown that overexpression of both the full-length HHR23A protein and a 179 amino acid

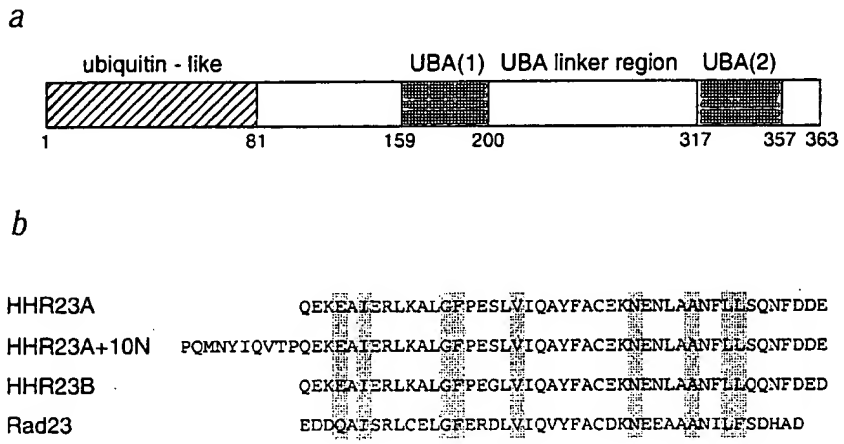


Fig. 1 ***a***, Sequence and domain structure of HHR23A. The different domains that have been identified by sequence analysis are indicated as shaded boxes. ***b***, Sequences of UBA domains used in this study. The amino acid positions that are conserved to more than 90% in all known UBA domains are highlighted.

C-terminal derivative, HHR23A(B213), which contains part of the UBA(1) domain, the UBA linker region and the entire UBA(2) domain results in significant alleviation of HIV-1 Vpr-induced cell cycle arrest. To further investigate the importance of the C-terminal UBA domain in cell cycle arrest, we constructed a plasmid pXCB213trunc, that expressed HHR23A(B213trunc), a C-terminal truncation containing a deletion of the entire UBA(2), domain to test in the alleviation assay, in which co-expression of Vpr and HHR23A(B213) results in partial alleviation of Vpr-induced G2 arrest (Fig. 2*a*, panels II and IV). In contrast, we found that cells co-expressing Vpr and HHR23A(B213trunc) were unable to alleviate HIV-1 Vpr-induced G2 arrest (Fig. 2*a*, panels IV and V). Thus, removal of the C-terminal 45 amino acids that comprise the UBA(2) domain of HHR23A(B213) was sufficient to abolish the alleviation of G2 arrest seen with the HHR23A(B213) although a stable truncated protein was expressed (Fig. 2*b*). These results confirm that the UBA(2) domain plays a central role in the alleviation of Vpr induced G2 arrest and that binding of Vpr to HHR23A is a crucial step in this process.

To obtain further information on the UBA(2) domain of HHR23A, the structure of a 45 amino acid peptide corresponding to residues 319–363 of HHR23A (Fig. 1) was determined by homonuclear NMR spectroscopy. The peptide was prepared by chemical synthesis and dissolved easily in the NMR buffer (50 mM potassium phosphate buffer at pH 6.5). The one-dimensional NMR spectrum of UBA(2) has a dispersed amide region indicative of a folded structure. The assignments and secondary structure were determined from analysis of two-dimensional DQF-COSY[13], TOCSY[14,15] and NOESY spectra[16] following established procedures[17]. The five N-terminal and C-terminal residues show no signs of any stable secondary structure; that is, there are no significant NOEs observable and the amide protons exchange rapidly with the solvent. This effect is also observed in the longer synthetic peptide with 10 amino acids added to the N-terminus (Fig. 1*b*, HHR23A+10N corresponding to residues 309–363 of HHR23A, data not shown) confirming that the residues 324–358 form the independently folding unit of the UBA motif. Qualitative analysis of the NOESY spectra (Fig. 3*a*) showed that UBA(2) forms a highly α-helical structure (Fig. 3*b*,*c*) as predicted by secondary structure prediction programs[18–20]. The three-dimensional structure of the molecule was determined based on NOE intensities extracted from 2D-NOESY spectra acquired with mixing times of 80 and 250 ms, using a combined distance geometry and simulated annealing approach[21,22]. A total of 50 structures were calculated, and the 10 structures with the lowest DYANA[23] target function value were selected for the final analysis. The structures are very well defined within the core region of the molecule (residues 5–42) (Fig. 4*a*). Input data and statistics of the resulting structures are summarized in Table 1.

The UBA domain folds into a compact three helix bundle with the helices oriented at angles of ~45° relative to the lon-

Table 1 Summary of input data and structure refinement parameters

Number of experimental restraints:	
Distance restraints from NOEs:	
Intra residue	82
Sequential	123
Medium range (i + 2 to i + 4)	86
Long range (> i + 4)	55
Number of NOEs per residue (residues 5–42)	9.4
Number of H-bond restraints (2 per bond)	34
Total number of experimental restraints:	380
R.m.s. deviations from experimental data:	
Number of refined structures	10
Average of DYANA target function	1.24 ± 0.07
Number of violations > 0.5 Å	none
Number of violations > 0.3 Å	6 (maximum 0.39 Å)
Average distance restraint violation	0.017 ± 0.001 Å
Ramachandran Analysis:	
Residues in favored regions	68.4%
Residues in additional allowed regions	26.8%
Residues in generously allowed regions	4.8%
Residues in disallowed regions	0.0%
Coordinate precision (Å):	
R.m.s. deviation of backbone atoms (residues 5–42)	1.24 ± 0.36 Å
R.m.s. deviation of all heavy atoms (residues 5–42)	2.22 ± 0.51 Å

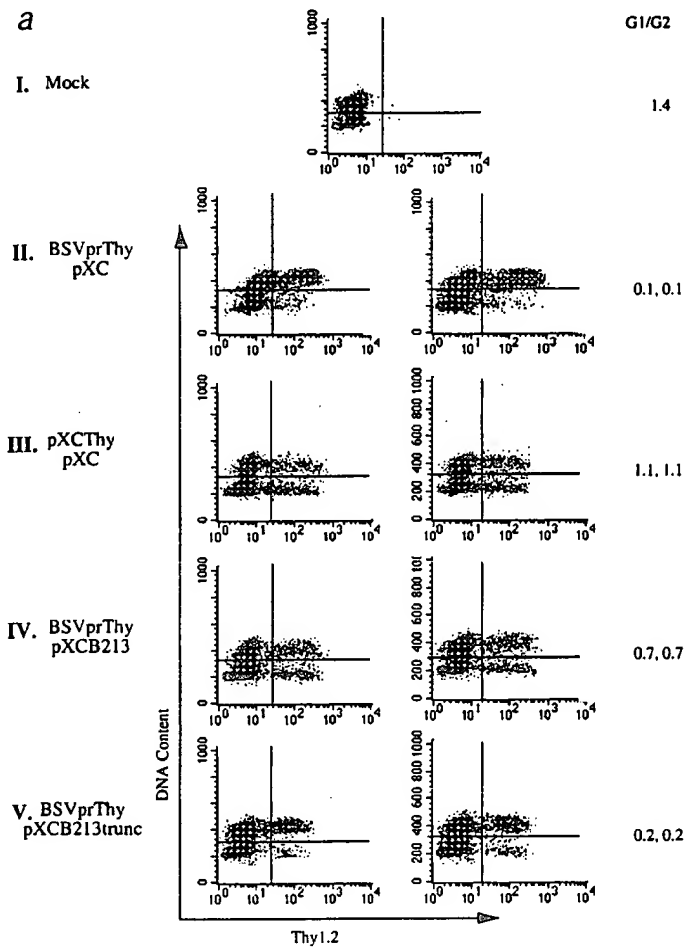


Fig. 2 Abrogation of alleviation of Vpr-induced cell cycle arrest by overexpression of a 134-residue C-terminal portion of HHR23A, HHR23A(B213trunc), that contains a deletion of the C-terminal UBA domain. *a*, Hela cells were co-transfected with either BSVprThy or pXCThy and a 20-fold molar excess of either pXC (3.2 μg), pXCB213 (4.0 μg), or pXCB213trunc (4.0 μg), and analyzed by flow cytometry as described[10]. Duplicate cotransfections are shown except for mock-transfected cells. Panel I, mock transfected cells; panel II, BSVprThy and pXC; panel III, pXCThy and pXC; panel IV, BSVprThy and pXCB213; panel V, BSVprThy and pXCB213trunc. *b*, Expression of HHR23A(B213trunc) in Hela cells. Hela cells were co-transfected with BSVprThy and pXCB213 or pXCB213trunc, and analyzed by Western blotting as described. Lanes 1 and 2 contain lysates from duplicate samples of cells co-transfected with BSVprThy and pXCB213 (shown in (*a*) panel IV), lanes 3 and 4 contain lysates from duplicate samples of cells co-transfected with BSVprThy and pXCB213trun c (shown in (*a*) panel V).

gitudinal axis of the molecule (Fig. 4*b*). The residues 324–330, 334–343 and 350–358 form three α-helices, α_1, α_2 and α_3 respectively (numbering scheme based on full size HHR23A). In addition to the structure of UBA(2) from HHR23A discussed here, the NOESY spectra of synthetic peptides corresponding to the sequences of the C-terminal UBA domains from HHR23B and yeast Rad23 (Fig. 1*b*) were analyzed. The assignments and the NOE distribution confirm that all these protein domains have the same secondary structure and overall fold. The pattern of conserved residues (Fig. 1*b*) in the different UBA domains can largely be explained by their participation in helix–helix interactions. The residues Gly 331, Phe 332 and Pro 333 form a sharp turn (which we call the GFP-loop) with Gly 331 and Phe 332 being highly conserved (Gly 331 to 100%, residue 332 is F or Y in 90% of the UBA sequences). The side chain of Phe 332 is sandwiched between the methyl groups of Leu 355 and Leu 356 connecting the GFP-loop between helices α_1 and α_2 to the C-terminal end of helix α_3 (Fig. 4*c*). The surface of the UBA domain has an uneven distribution of charged and hydrophobic amino acids (Fig. 4*d,e*). This large hydrophobic area on the bottom of the helix bundle in the vicinity of the GFP-loop is the most likely interface for protein–protein interactions.

DNA damage and Vpr action both lead to cell cycle arrest at the same checkpoint which suggests that the same pathways may be used. The binding of Vpr to UBA(2) of HHR23A seems to be a crucial step in this process (this paper and ref. 10). The N-terminal ubiquitin-like domain of yeast Rad23 has recently been shown to interact with the proteosome[24]. A similar role for the HHR23A and B proteins is likely based on the extensive homology between them and Rad23 with respect to sequence and function. The linker region of HHR23B has been shown to be involved in XPC (Xeroderma pigmentosum group C) binding[25]. Thus the Rad23 homologs provide a link between the ubiquitin/proteosome and DNA excision repair pathways and are potentially involved in the regulation of the assembly of the repair complex. In this context it is interesting that native Rad23 is ubiquitinated and degraded during the transition between the G1 and S phases of the cell cycle[24]. It is possible that binding of Vpr disrupts this link between cell cycle control and DNA repair through the interaction with UBA(2). The structure of the UBA(2) domain of HHR23A shown here provides a rational basis for designing mutant proteins to test and further define the Vpr binding site on HHR23A. Because of the multiple functions of Vpr important for HIV-1 replication as well as pathogenesis, the development of therapeutics

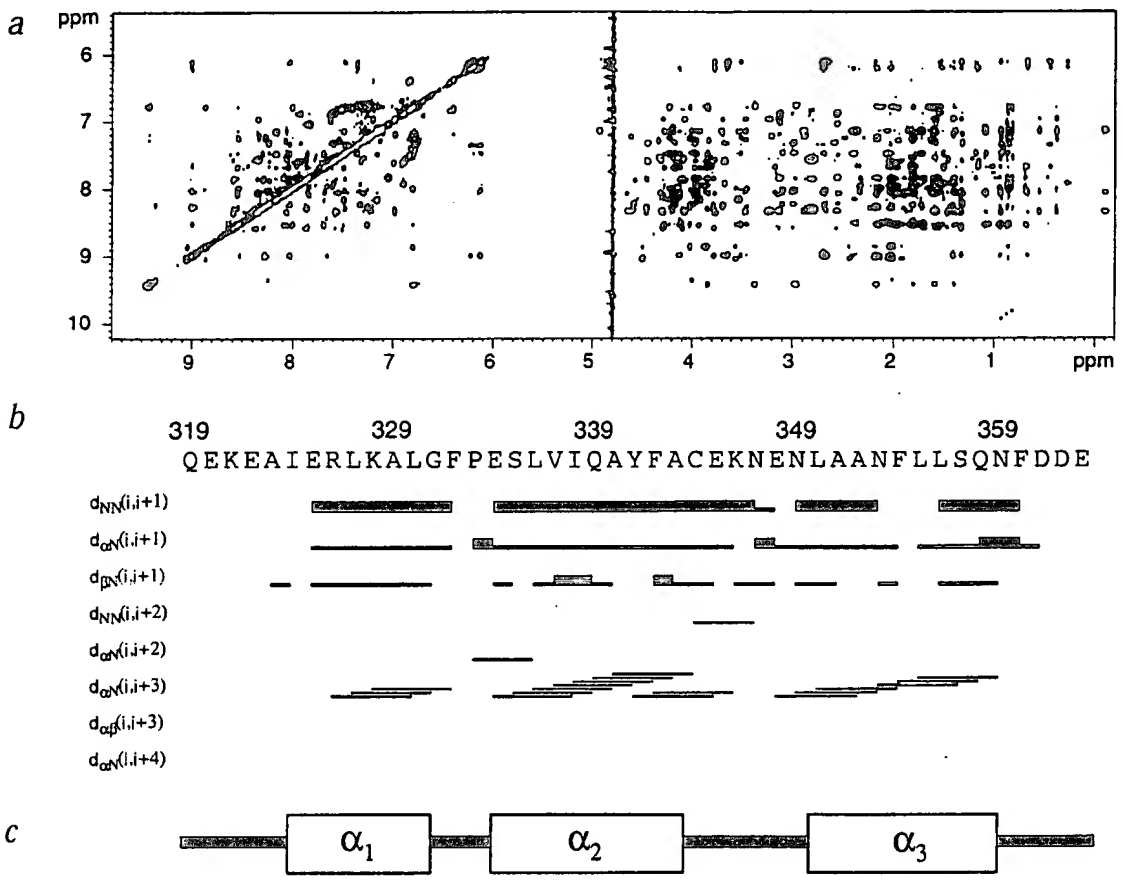


Fig. 3 *a*, 300 ms 2D WATERGATE-NOESY spectrum of the HHR23A UBA domain 2. The spectrum was acquired at 293 K on a sample containing 1.0 mM synthetic peptide in 450 µl 50 mM potassium phosphate buffer at pH 6.5. A total of 600 t_1 increments of 64 scans each were acquired in States-TPPI mode. The spectrum was processed with 2,048*2,048 real points after multiplication with a squared sine bell function shifted by $\pi/3$ and $\pi/2$ in F_1 and F_2 respectively. *b*, Summary of NOE contacts for HHR23A UBA(2). The numbering scheme is that of the full size HHR23A. *c*, NMR derived secondary structure of the UBA domains of HHR23A.

directed against Vpr as a target would provide another line of attack against HIV-1 disease. The understanding achieved here regarding a critical domain that interacts with Vpr is a first step towards meeting this goal.

Methods

Cells. HeLa cells were grown in Dulbecco's modified Eagle's medium (D-MEM) supplemented with calf serum (10%), penicillin (100 U ml^{-1}) and streptomycin (100 µg ml^{-1}). At a time of 18–24 h prior to transfection cells were plated in six-well plates at a density of 2×10^5 per well. HeLa cells were cotransfected with either BSVprThy or pXCThy and ~20-fold molar excess of either pXC, pXCB213, or pXCB213trunc using Lipofectin (Gibco BRL) as directed by the manufacturer. At 48 h after transfection cells were analyzed by flow cytometry as described below.

Mammalian expression constructs. The plasmids BSVprThy, pXCThy and pXC have been described[10]. pXCB213trunc contains a truncation of the C-terminal UBA domain of the HHR23A(B213) fragment, designated UBA(2), and was constructed by insertion of PCR amplified HHR23A sequences (nucleotides 589–990) into the *Xho*I and *Mlu*I sites of the expression plasmid pXC. The sequences of the primers (provided by DNA Technology Group, Amgen, Boulder, CO) used for PCR amplification were: 5'-CCGCTCGAGATG**GACTACAAAGACGATGACGATAAA**AGAGCCAGCTACAACAAC-3' (forward primer) and 5'-CCGACGCGTTCA- TCACGGCGTCACCTGGATGTAGT-3' (reverse primer). The forward primer contains an *Xho*I restriction site, a translation initiation site and an M2 FLAG epitope tag (bold). The reverse primer contains an *Mlu*I site at its 5' end. The DNA sequence of the *Xho*I–*Mlu*I fragment in pXCB213trunc was confirmed by sequence analysis. The HHR23A(B213) expression plasmid pXCB213 encodes the C-terminal portion of HHR23A (residues 185–363). The HHR23A(B213trunc) expression plasmid pXCB213trunc encodes a truncated C-terminal portion of HHR23A (residues 185–318) and contains a deletion of the C-terminal UBA(2) domain. The BSVprThy expression plasmid encodes the HIV-1 Vpr protein (96 residues) and the murine Thy1.2 protein (162 residues), a cell surface marker that allows us to distinguish successfully transfected cells. The pXCThy expression plasmid encodes the murine Thy1.2 protein as described above. The pXC expression plasmid does not encode any proteins and is used as a control.

Flow cytometry. At 48 h after transfection, cells were stained with a monoclonal antibody to the Murine Thy1.2 cell-surface protein directly conjugated with fluorescein isothiocyanate (FITC) and propidium iodide (PI) and analyzed by flow cytometry using

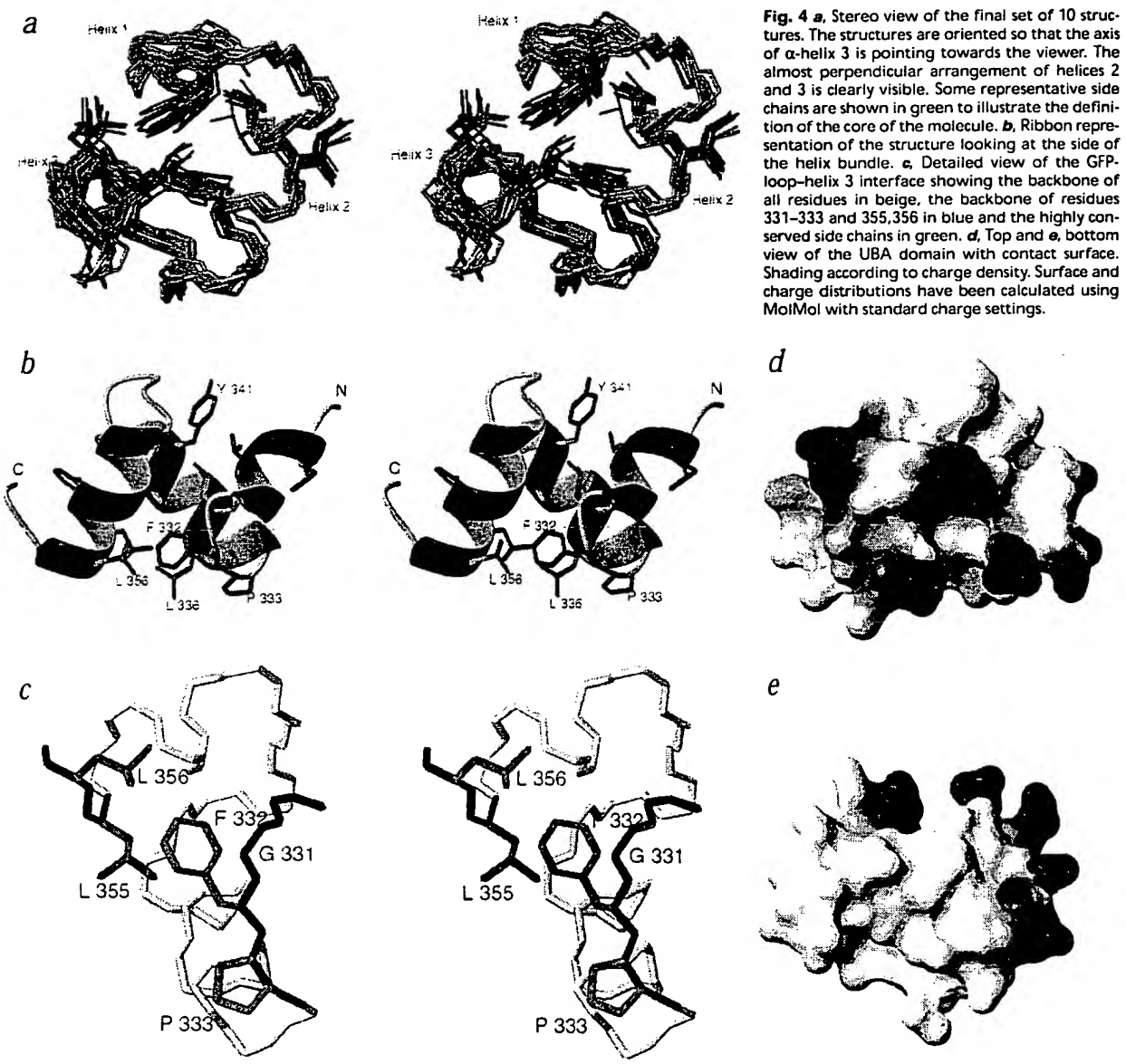


Fig. 4 ***a***, Stereo view of the final set of 10 structures. The structures are oriented so that the axis of α-helix 3 is pointing towards the viewer. The almost perpendicular arrangement of helices 2 and 3 is clearly visible. Some representative side chains are shown in green to illustrate the definition of the core of the molecule. ***b***, Ribbon representation of the structure looking at the side of the helix bundle. ***c***, Detailed view of the GFP-loop–helix 3 interface showing the backbone of all residues in beige, the backbone of residues 331–333 and 355,356 in blue and the highly conserved side chains in green. ***d***, Top and ***e***, bottom view of the UBA domain with contact surface. Shading according to charge density. Surface and charge distributions have been calculated using MolMol with standard charge settings.

the Lysis II software (Becton Dickson) as described[10]. Gates were set to exclude clumps and cell debris and on the Thy1.2$^-$ and Thy1.2$^+$ populations. The Thy1.2$^-$ and Thy1.2$^+$ populations were distinguished by comparing the staining pattern of Thy1.2 transfected populations to that of a mock-transfected population of cells. A total of 800–2,000 Thy1.2$^+$ events were acquired for each sample except in the case of mock infected cells, where 10,000 Thy1.2$^-$ events were acquired. The (G1/G2+M) ratios presented are those calculated from the Thy1.2$^+$ cells except in the case of mock transfected cells, where the ratio represents that of the Thy1.2$^-$ population. The Thy1.2-FITC fluorescence intensity is plotted on the X axis and the fluorescence intensity of PI is plotted on the Y axis.

Western blot analysis. Cell lysates from transfected Hela cells were prepared as described[10]. Proteins were separated by SDS PAGE and electro-blotted to nitrocellulose membranes. Blots were incubated with a mouse monoclonal antibody specific for the M2 FLAG (Eastman Kodak Co.) followed by incubation with a horseradish peroxidase-linked sheep anti-mouse immunoglobulin (Amersham Life Science Inc.). Antibody binding was detected using the ECL detection kit (Amersham Life Science Inc.).

Synthesis of peptides and sample preparation. All peptides were synthesized by the Fmoc (fluorenylmethoxycarbonyl)/t-butyl based solid phase peptide chemistry method on an ABI 431A (Perkin Elmer) peptide synthesizer utilizing a single coupling program to carry out the chain assembly. The crude peptide was dissolved in 8 M guanidine-HCl, and purified to >95% homogeneity by preparative reverse-phase HPLC on a Vydac C_{18} (2.5 cm × 25 cm) column with linear gradients consisting of 0.1%TFA (v/v) in water and 0.05%TFA (v/v) in acetonitrile. The composition and purity of each peptide was assessed using electro-spray ionization (ESI) mass spectrometry and amino acid analysis. NMR samples were prepared by dissolving the purified, lyophilized peptide in the NMR buffer (50 mM potassium phosphate buffer at pH 6.5 in 99.9996% D_2O or 90% H_2O/10% D_2O). Undissolved peptide was spun down in a microcentrifuge prior to transferring the samples into standard 5 mm NMR tubes. The resulting samples were between 0.8 and 1.0 mM in peptide in 450 µl solution.

NMR spectroscopy. All spectra were acquired on Bruker Avance DRX spectrometers at 500 and 600 MHz. Water suppression was achieved using presaturation or WATERGATE[26] sequences. Two-dimensional DQF-COSY[13], TOCSY[14,15] and NOESY[16] spectra were measured at 293 K in 100% D_2O and 90% H_2O/10% D_2O at 293 K. The TOCSY datasets were acquired using the CITY[27] or MLEV-17[28] mixing sequences with mixing times of 50 and 80 ms, for NOESY spectra mixing times of 80, 150, 250 and 300 ms were applied. The spectra were acquired with data matrices of 300*1024 complex points in F_1 and F_2, respectively. For the DQF-COSY 800 points were acquired in F_1. In addition a ^1H-^{13}C correlation at natural abundance was acquired using a sensitivity enhanced, gradient selected HSQC pulse sequence[29]. For this data set 160 scans per t_1 increment were acquired with 40*1024 complex points in F_1 and F_2 respectively. All spectra were processed using squared sine bell window functions that were shifted by $\pi/2$ and $\pi/3$ in F_1 and F_2 respectively. The final data matrices for all data sets were 1024*1024 real points.

Data analysis and structure calculations. The data analysis and the assignment of proton resonances were facilitated by the use of the software package XEasy. The structure calculations were done with the program DYANA[23]. A total of 50 structures were generated using the standard DYANA protocol; the 10 structures with the lowest value for the target function were used for the final analysis using the programs MolMol[30] and InsightII (Biosym Inc.). Protein secondary structure prediction was performed using the EMBL PHD mail server[31].

Coordinates. The coordinates of the ten final structures of the HHR23A C-terminal UBA domain have been deposited in the Protein Data Bank (accession number 1UBA).

Acknowledgment
This work was supported by NIH grants to J.F., I.S.Y.C. and E.S.W.-W.

Thorsten Dieckmann[1], Elizabeth S. Withers-Ward[2], Mark A. Jarosinski[3], Chuan-Fa Liu[3], Irvin S. Y. Chen[2] and Juli Feigon[1]

[1]Deptartment of Chemistry and Biochemistry and the Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90095, USA. [2]Departments of Microbiology and Immunology and Medicine, and UCLA AIDS Institute, Los Angeles, CA 90095, and [3]Amgen Inc., Department of DNA and Peptide Chemistry, Boulder, CO 80301.

Correspondence should be addressed to J.F. *e-mail: feigon@mbi.ucla.edu*

Received 2 June, 1998; accepted 11 September, 1998.

1. Huang, L.M. & Jeang, K.T. In *Human retroviuses and AIDS* (eds Myers, G., Hahn, B., Mellors, J. & Henderson, L.) III 3-9 (Los Alamos National Laboratory, Los Alamos; 1995).
2. Vodicka, M.A., Koepp, D.M., Silver, P.A. & Emerman, M. *Genes Dev.* **12**, 175–185 (1998).
3. Jowett, J.B.M. *et al. J. Virol.* **69**, 6304–6313 (1995).
4. He, J. *et al. J. Virol.* **69**, 6705–6711 (1995).
5. Re, F., Braaten, D., Franke, E.K. & Luban, J. *J. Virol.* **69**, 6859–6864 (1995).
6. Rogel, M., Wu, L.I. & Emerman, M. *J. Virol.* **69**, 882–888 (1995).
7. Poon, B., Grovit-Ferbas, K., Stewart, S.A. & Chen, I.S.Y. *Science* **281** 266–269 (1998).
8. Stewart, S.A., Poon, B., Jowett, J.B.M. & Chen, I.S.Y. *J. Virol.* **71**, 5579–5592 (1997).
9. Poon, B. *et al. J. Virol.* **71**, 3961–3971 (1997).
10. Withers-Ward, E. *et al. J. Virol.* **7**, 9732–9742 (1997).
11. Watkins, J.F., Sung, P., Prakash, L. & Prakash, S. *Mol. Cell. Biol.* **13**, 7757–7765 (1993).
12. Hofman, K. & Bucher, P. *TIBS Biol.* **21**, 172–173 (1996).
13. Rance, M. *et al. Biochem. Biophys. Res. Comm.* **117**, 479–485 (1983).
14. Bax, A. & Davis, D.G. *J. Magn. Reson.* **65**, 355–360 (1985).
15. Davis, D.G. & Bax, A. *J. Am. Chem. Soc.* **107**, 2820–2821 (1985).
16. Macura, S., Huang, Y., Suter, D. & Ernst, R.R. *J. Magn. Reson.* **43**, 259–281 (1980).
17. Wüthrich, K. *NMR of proteins and nucleic acids.* 292 (John Wiley & Sons, New York, New York; 1986).
18. Rost, B. & Sander, C. *Proc. Natl. Acad. Sci. USA* **90**, 7558–7562 (1993).
19. Rost, B. & Sander, C. *J. Mol. Biol.* **232**, 584–599 (1993).
20. Rost, B., Sander, C. & Schneider, R. *J. Mol. Biol.* **235**, 13–26 (1994).
21. Brünger, A.T. & Nilges, M. *Quart. Rev. Biophys.* **26**, 49–125 (1993).
22. Nilges, M. *Curr. Opin. Struct. Biol.* **6**, 617–623 (1996).
23. Guntert, P., Mumenthaler, C. & Wüthrich, K. *J. Mol. Biol.* **273**, 283–298 (1997).
24. Schauber, C. *et al. Nature* **391**, 715–718 (1998).
25. Masutani, C. *et al. Mol. Cell. Biol.* **17**, 6915–6923 (1997).
26. Piotto, M., Saudek, V. & Sklenář, V. *J. Biomol. NMR* **2**, 661–665 (1992).
27. Kadkhodaei, M., Hwang, T.L., Tang, J. & Shaka, A.J. *J. Magn. Reson. Ser. A* **105**, 104–107 (1993).
28. Levitt, M.H., Freeman, R. & Frenkiel, T. *J. Magn. Reson.* **47**, 328–330 (1982).
29. Kay, L.E., Keifer, P. & Saarinen, T. *J. Am. Chem. Soc.* **114**, 10663–10665 (1992).
30. Koradi, R., Billeter, M. & Wüthrich, K. *J. Mol. Graphics* **14**, 51+ (1996).
31. Rost, B., Sander, C. & Schneider, R. *Comput. Appl. Biosci.* **10**, 53–60 (1994).

Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP

The Rho-related small GTP-binding protein Cdc42 has a low intrinsic GTPase activity that is significantly enhanced by its specific GTPase-activating protein, Cdc42GAP. In this report, we present the tertiary structure for the aluminum fluoride-promoted complex between Cdc42 and a catalytically active domain of Cdc42GAP as well as the complex between Cdc42 and the catalytically compromised Cdc42GAP(R305A) mutant. These structures, which mimic the transition state for the GTP hydrolytic reaction, show the presence of an AlF_3 molecule, as was seen for the corresponding Ras–p120RasGAP complex, but in contrast to what has been reported for the Rho–Cdc42GAP complex or for heterotrimeric G protein α subunits, where AlF_4^- was observed. The Cdc42GAP stabilizes both the switch I and switch II domains of Cdc42 and contributes a highly conserved arginine (Arg 305) to the active site. Comparison of the structures for the wild type and mutant Cdc42GAP complexes provides important insights into the GAP-catalyzed GTP hydrolytic reaction.

GTPase-activating proteins (GAPs) are critical regulators of all signaling pathways mediated by members of the Ras superfamily. GAPs serve to attenuate these signals by accelerating the hydrolysis of GTP to GDP. Recent biochemical and structural studies demonstrating the ability of aluminum fluoride to act as a transition state analog in the hydrolytic pathway of low molecular weight GTP-binding proteins has provided a new tool to investigate the mechanism of GTP hydrolysis and the role of GAPs in accelerating this reaction[1]. Previous studies from our laboratory have demonstrated the ability of aluminum fluoride to induce stable complex formation between the GDP-bound form of Cdc42 and its Cdc42GAP[2]. Here, we report the crystal structures of the aluminum fluoride-promoted transition state complex between Cdc42 and the wild type Cdc42GAP (sometimes referred to as RhoGAP[3]), as well as the complex between Cdc42 and the catalytically compromised Cdc42GAP(R305A) mutant and discuss the mechanistic implications of these structures for GAP-catalyzed GTP hydrolysis.

bioSCOUT®

Projects | Analyses | Admin | Alert | SRS | Tutorial | Help | Features

Summary

Searched query 0964rad23.sid2 against PFAM database.

Hit	Score	Expect	Description	Q from	Q to	Method
☐ pfam|hmm|ubiquitin . alignment	89.3	7.8e-23	Ubiquitin family	1	78	HMMPFAM
☐ pfam|hmm|UBA . alignment	52.5	9.5e-12	UBA/TS-N domain	159	201	HMMPFAM
☐ pfam|hmm|UBA . alignment	33.6	4.5e-06	UBA/TS-N domain	357	396	HMMPFAM
☐ pfam|hmm|UPF0184 . alignment	-24.6	7.2	Uncharacterised protein family (UPF0184) -	244	337	HMMPFAM

New Task | Rename Sequences

Please report problems and feedback concerning bioSCOUT through the support interface.

Alignment: 0964rad23.sid2 - pfam|hmm|ubiquitin

HMMPFAM - alignment of 0964rad23.sid2 against pfam|hmm|ubiquitin

Ubiquitin family

- This hit is scoring at : 89.3
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)

```
Q:       1 MKLNVKTLKGT-------NFEIEASPDASVADVKRIIETTQGqstYRADQQMLIYQGKIL
           MK: VKTL.G.        ...:E..P. :V.::K. IE..:G    . .:QQ.LIY.GK:L
H:       1 mkifVKtldgkgkvtdgstitlevepsdTVlelKekIedkeg...ippeqQrLiyaGkvL

            -KDETTLESNGVAENSFLVIMLSKAK      78
             :D:TTL.. .:.E.SF.: :: :.:
            reDdttLaeYniqegsftlhlvlrlr      83
```

Legend of Alignment

`:` positive score

`.` score between -2 and 0

Please report problems and feedback concerning bioSCOUT through the support interface.

Protein families database of alignments and HMMs

ubiquitin

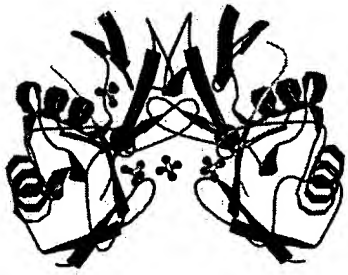


Figure 1: 1ndd
Signaling protein
Structure of nedd8

Accession number: **PF00240**

Ubiquitin family

This family contains a number of ubiquitin-like proteins: SUMO (smt3 homologue) (see ULP1_YEAST), Nedd8 (see NED8_MOUSE), Elongin B (see Q15370), Rub1 (see Q9SHE7).

NEW! This family forms **structural complexes** with other Pfam families, to view them click here

---

INTERPRO description (entry IPR000626)

Ubiquitin [MEDLINE:91274342], PUB00000768, PUB00000768 is a protein of seventy six amino acid residues, found in alleukaryotic cells and whose sequence is extremely well conserved from protozoan to vertebrates. It is widely known as a post-translational tag used to signal a protein's hydrolytic destruction. Other functions for ubiquitin, depend on its differential internal isopeptide linkages. In addition, several ubiquitin-like proteins have been discovered from genome-sequencing efforts, other structural studies, and genetic screens. These new data show that proteins with the ubiquitin domain are adaptable, transposable genetic elements, which have been appended to other genes and utilized for many different cellular functions, depending on the ubiquitin-like protein's identity, subcellular location, and method of covalent attachment. The post-translational ligation of proteins to members of the ubiquitin superfamily can signal many different fates for the target protein [Larsen and Wang (2002) J. Proteome Res. 1, 411-419.]

Ubiquitin is a globular protein, the last four C-terminal residues (Leu-Arg-Gly-Gly) extending from the compact structure to form a 'tail' important for its function. The latter is mediated by the covalent conjugation of ubiquitin to target proteins, by an isopeptide linkage between the C-terminal glycine and the epsilon amino group of lysine residues in the target proteins.

In most species, there are many genes coding for ubiquitin. However they can be classified into two classes. The first class produces polyubiquitin molecules consisting of exact head to tail repeats of ubiquitin. The number of repeats is variable. In the majority of polyubiquitin precursors, there is a final amino-acid after the last repeat. The second class of genes produces precursor proteins consisting of a single copy of ubiquitin fused to a C-terminal extension protein (CEP). There are two types of CEP proteins and both seem to be ribosomal proteins. There are a number of proteins which are evolutionary related to ubiquitin, including ubiquitin-like proteins from baculoviruses; mammalian proteins GDX, FAU and RAD23-related proteins; human spliceosome associated protein 114, proteins BAT3 and CKAP1/TFCB, and ubiquitin-like proteins SMT3A, SMT3C and SMT3B; yeast proteins RAD23, DK2 and SMT3; and C. elegans SMT3 and ubl-1 proteins.

For additional annotation, see the PROSITE document PDOC00271 [Expasy | SRS-UK | SRS-USA]

Alignment	Domain organisation
◉ Seed (95) ○ Full (1322) Format Coloured alignment Get alignment Further alignment options here Help relating to Pfam alignments here	◉ Seed (95) ○ Full (1322) As a Graphic: Zoom 0.5 pixels/aa. View Graphic As a Tree: ☐ Bootstrap tree NIFAS Applet To find out about the NIFAS tree-viewer, click here
Species Distribution	**Phylogenetic tree**
NEW! View alignments & domain organisation by species Tree depth : Show all levels View Species Tree	◉ Seed (95) ○ Full (1322) Download tree ATV Applet The trees were generated using Quicktree To find out more about ATV phylogenetic tree-viewer click here

Database References	
PDB You can find out how to set up Rasmol here	1a5r ; 20; 95; PDB 2 Pfam Rasmol (unix) CATH-PDBSUM SCOP-UK
PRINTS	PR00348
PROSITE	PDOC00271 [Expasy | SRS-UK | SRS-USA]
HOMSTRAD	UBQ
PFAMB	PB013905 PB018771 PB024244 PB038815 PB039385 PB047048 PB073 PB074986 PB075222 PB075546 PB076014
SYSTERS	ubiquitin
PANDIT	ubiquitin

Literature References
1. **Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes.** **Whitby FG, Xia G, Pickart CM, Hill CP;** **J Biol Chem 1998;273:34983-34991.**
2. **Structure determination of the small ubiquitin-related modifier SUMO-1.** **Bayer P, Arndt A, Metzger S, Mahajan R, Melchior F, Jaenicke R,**

Pfam specific information	
Author of entry	Finn RD, Griffiths-Jones SR
Type definition	Domain
Alignment method of seed	Clustalw
Source of seed members	Prosite

HMMER build information		
	Pfam_ls [Download HMM]	**Pfam_fs** [Download HMM]
Gathering cutoff	25.5 25.5;	11.2 11.2

Becker J;
J Mol Biol 1998;280:275-286.

3.
Structure of tetraubiquitin shows how multiubiquitin chains can be formed.
Cook WJ, Jeffrey LC, Kasperek E, Pickart CM;
J Mol Biol 1994;236:601-609.

4.
Structure of ubiquitin refined at 1.8 A resolution.
Vijay-Kumar S, Bugg CE, Cook WJ;
J Mol Biol 1987;194:531-544.

Trusted cutoff	26.2 26.1;	11.2 11.2
Noise cutoff	25.2 25.2;	11.1 11.1
Build method of HMM	hmmbuild -F HMM_ls SEED hmmcalibrate --seed 0 HMM_ls	hmmbuild -f -F HMM_fs SEED hmmcalibrate --seed 0 HMM_fs

For help on making stable links to this page click here

Comments or questions on the site? Send a mail to **pfam@sanger.ac.uk**

Alignment: 0964rad23.sid2 - pfam|hmm|UBA

HMMPFAM - alignment of 0964rad23.sid2 against pfam|hmm|UBA

UBA/TS-N domain

- This hit is scoring at : 52.5
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)

```
Q:     159 NLEQTIQQILDMGGGTWERDTVVRALRAAYNN--PERAIDYLYSG     201
             E:.::Q: :MGGG.:.R:.. :ALRA. NN   ERA:::L..
H:       1 ideekvkqLreMGGGvfsreeakkALratnnnsDverAvewLleh      45
```

HMMPFAM - alignment of 0964rad23.sid2 against pfam|hmm|UBA

UBA/TS-N domain

- This hit is scoring at : 33.6
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)

```
Q:     357 EEREAIQRLEGMG---FNRELVLEVFFACNK--DEELTANYLLDH     396
            :.E.:::L. MG   F:RE . :.. A.N.  D E ...:LL:H
H:       1 ideekvkqLreMGGGvfsreeakkALratnnnsDverAvewLleh      45
```

Legend of Alignment

```
:  positive score

.  score between -2 and 0
```

Please report problems and feedback concerning bioSCOUT through the support interface.

Protein families database of alignments and HMMs

Home | Keyword Search | Protein Search | Browse Pfam | DNA Search | Taxonomy | ftp | H

UBA

Figure 1: 1aip
Complex of two elongation factors
Ef-tu ef-ts complex from thermus thermophilus

Accession number: **PF00627**

UBA/TS-N domain

This small domain is composed of three alpha helices. This family includes the previously defined UBA and TS-N domains. The UBA-domain (ubiquitin associated domain) is a novel sequence motif found in several proteins having connections to ubiquitin and the ubiquitination pathway. The structure of the UBA domain consists of a compact three helix bundle [1]. This domain is found at the N terminus of EF-TS hence the name TS-N. The structure of EF-TS is known and this domain is implicated in its interaction with EF-TU [2]. The domain has been found in non EF-TS proteins such as alpha-NAC P70670 and MJ0280 Y280_METJA [1].

NEW! This family forms **structural complexes** with other Pfam families, to view them click here

---

INTERPRO description (entry IPR000449)

UBA domains are a commonly occurring sequence motif of approximately 45 amino acid residues that are found in diverse proteins involved in the ubiquitin/proteasome pathway, DNA excision-repair, and cell signaling via protein kinases [MEDLINE:97025177]. The human homologue of yeast Rad23A is one example of a nucleotide excision-repair protein that contains both an internal and a C-terminal UBA domain.

The solution structure of human Rad23A UBA(2) showed that the domain forms a compact three-helix bundle [MEDLINE:99061330]. Comparison of the structures of UBA(1) and UBA(2) reveals that both form very similar folds and have a conserved large hydrophobic surface patch which may be a common protein-interacting surface present in diverse UBA domains. Evidence that ubiquitin binds to UBA domains leads to the prediction that the hydrophobic surface patch of UBA domains interacts with the hydrophobic surface on the five-stranded β-sheet of ubiquitin [MEDLINE:22075341].

For additional annotation, see the PROSITE document PDOC50015 [Expasy | SRS-UK | SRS-USA]

Alignment	Domain organisation
Seed (162) Full (473)	Seed (162) Full (473) Context (51)
Format Coloured alignment	As a Graphic / As a Tree
Get alignment	Zoom 0.5 pixels/aa. Bootstrap tree
Further alignment options here	View Graphic / NIFAS Applet
Help relating to Pfam alignments here	To find out about the NIFAS tree-viewer, click here

Species Distribution	Phylogenetic tree
NEW! View alignments & domain organisation by species Tree depth : Show all levels View Species Tree	Seed (162) Full (473) Download tree ATV Applet The trees were generated using Quicktree To find out more about ATV phylogenetic tree-viewer click here

Database References	
PDB You can find out how to set up Rasmol here	1aip C; 2; 42; PDB 2 Pfam Rasmol (unix) CATH-PDBSUM SCOP-UK
URL	http://www.isrec.isb-sib.ch/profile/isrec96/poster.html
PROSITE	PDOC50015 [Expasy | SRS-UK | SRS-USA]
PROSITE_PROFILE	PS50030
COGS	COG0264 COG1308
HOMSTRAD	EF_TS
PFAMB	PB005809 PB015819 PB016122 PB038705 PB073598 PB073788 PB074(
SYSTERS	UBA
PANDIT	UBA

Literature References
1. The structure of the Escherichia coli EF-Tu.EF-Ts complex at 2.5 A resolution. **Kawashima T, Berthet-Colominas C, Wulff M, Cusack S, Leberman R; Nature 1996;379:511-518.**
2. **Alignment of TS-N domain is in figure 7c.** Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. **Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV; Genome Res 1999;9:608-628.**
3. Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr.

Pfam specific information	
Author of entry	Bateman A
Type definition	Domain
Alignment method of seed	Manual
Source of seed members	Bateman A

HMMER build information		
	Pfam_ls [Download HMM]	**Pfam_fs** [Download HMM]
Gathering cutoff	17.3 17.3;	19.0 19.0
Trusted cutoff	17.3 17.3;	19.0 19.0
Noise cutoff	17.2 17.2;	18.9 18.9
Build method of HMM	hmmbuild -F HMM_ls SEED hmmcalibrate --seed 0 HMM_ls	hmmbuild -f -F HMM_fs SEED hmmcalibrate --seed 0 HMM_fs

Dieckmann T, Withers-Ward ES, Jarosinski MA, Liu CF, Chen IS, Feigon J;
Nat Struct Biol 1998;5:1042-1047.

For help on making stable links to this page click here

Comments or questions on the site? Send a mail to **pfam@sanger.ac.uk**

Alignment: 0964rad23.sid2 - pfam|hmm|UPF0184

HMMPFAM - alignment of 0964rad23.sid2 against pfam|hmm|UPF0184

Uncharacterised protein family (UPF0184) -

- This hit is scoring at : -24.6
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)

```
Q:     244 -AGPNANPlnlfpqGVPSGGSNPGVVPGAGSGALDAL-RQLPQFQALLQLVQANPQILQP
            :GPN.:.         .  ...    .  G  ..:A: ..L.Q..:.L..::.. . L..
H:       1 MSGpNgdd......psSvedggntnedefglqEyeaiNsmLDqinsaLDalEernDdlms

            MLQELGKQNPQILRLIQENqaeflrLVNESPEGGPG     337
            .L:EL :.N QI   ..E.      L N.:PE.G..
            kLrELLeSNRqiRlefaeq......ldnnAPesgsd      84
```

Legend of Alignment

```
:  positive score

.  score between -2 and 0
```

Please report problems and feedback concerning bioSCOUT through the support interface.

bioSCOUT®

Projects | Analyses | Admin | Alert | SRS | Tutorial | Help

Analysis Browser: Level Up

Report for 0964rad23.sid4 (Protein) Update

Description 0964_rad23_sid4 Edit

Function

RAD23 protein, isoform II

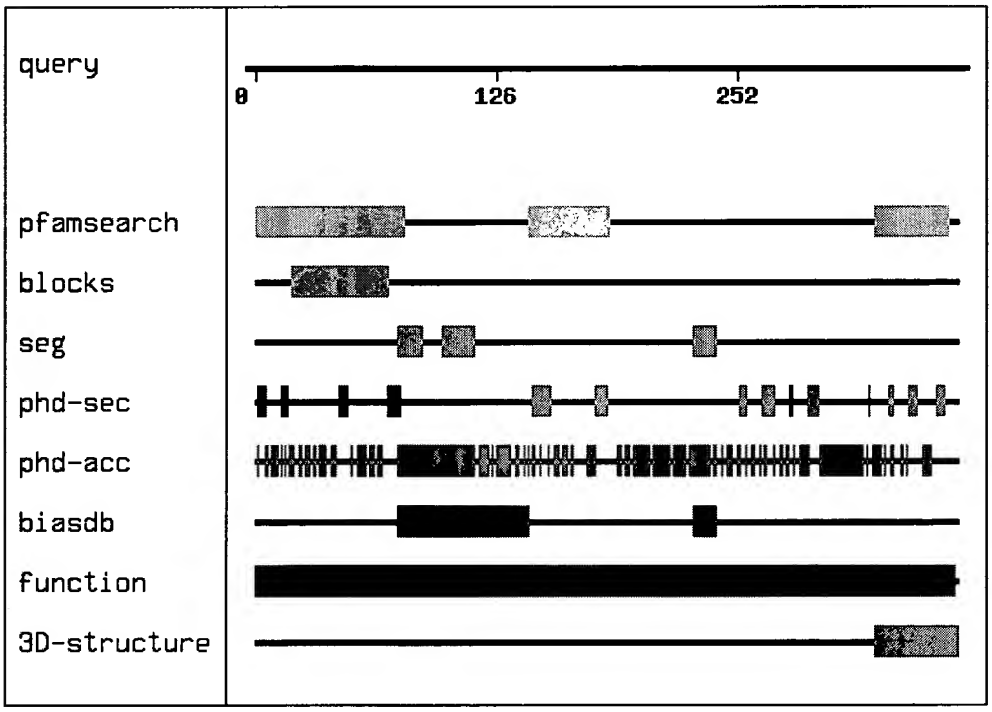
Direct assignment of functionality by homology to trembl|Y12014|DCRAD23II_1

in region **1** to **366** for overall length of **21** (99% of query, 1742% of hit, see the alignment).

Functional class Replication

Extracted keywords DNA damage, Nuclear protein, DNA repair

Features Summary



Homologies

All BLAST hits

Protein **70 clear** homologs All protein BLAST hits

	ESTs	**180** homologs	All EST BLAST hits
	Patents	**177** homologs	All patent hits
General	**Gene name**		
	Molecular weight	39.87 kD	
	Sequence length	368	
	Isoelectric point	4.41	
	Predicted cellular localisation (PHD and PreLoc)	cytoplasmic (64.8 %)	
	Identical sequence segments in:	embl|AI901927|AI901927 genbank|AI901927|AI901927	
3D Structure	**3D structure inferred by clear homology from residues 324 to 368 in 1F4I-A**		
	View	alignment	
	pdb|1F4I|1F4I-A	structure	
Expression	**Expression of this gene is reported for**		
	Organ Category	other species	
	Development Stage	tassel length from 0.1 to 2.5 cm	
	Tissue Classification	normal tissue	
Phylogeny	**Distribution**	32 species extracted from 189 homologous sequences.	Species
	Taxa	Chordata, Eukaryotae, Fungi, Planta	
	Model organisms	*Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Homo sapiens, Mus musculus, Saccharomyces cerevisiae*	
Features	**Low complexity region**	from **75** to **88,** from **98** to **115,** from **230** to **241** detected by [seg]	
	P-rich region	from **94** to **129** detected by [biasdb]	
	S-rich region	from **75** to **144** detected by [biasdb]	
	G-rich region	from **230** to **241** detected by [biasdb]	
	No significant hits	[Coils] [Phd-tm]	

	detected by	
Patterns	**UBA/TS-N domain region**	from residue **324** to **363.** Source: [pfamsearch] . Quality: (E=2e-09)
		from **144** to **186.** Source: [pfamsearch] . Quality: (E=2.7e-11)
	Ubiquitin family region	from residue **1** to **78.** Source: [pfamsearch] . Quality: (E=2.3e-23)
	No significant hits found in	[prosite database] [blocks database]

Comment No comment section. Edit

Completed Tasks

Start Time	User	Comment	Output	Interactiv
03.04.2003, 12:46:38	dressvm		bioSCOUT_default details...	

Permissions Edit

Alert Jobs New

Please report problems and feedback concerning bioSCOUT through the support interface.

This entry is from: TREMBL:DCRAD23II_1


```
ID   DCRAD23II_1 standard; PRT; 379 AA.
AC   Y12014;
DR   EMBL; Y12014; DCRAD23II.
DE   product: "RAD23 protein, isoform II";
DE   Daucus carota mRNA for RAD23 protein, isoform II
OS   Daucus carota (carrot)
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheopl
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; A
OC   euasterids II; Apiales; Apiaceae; Daucus.
FT   CDS             40..1179
FT                   /db_xref="SPTREMBL:O03991"
FT                   /product="RAD23 protein, isoform II"
FT                   /function="assembly factor of the complex for
FT                   excision repair of V-damaged DNA"
FT                   /protein_id="CAA72742.1"
FT                   /translation="MKLTVKTLKGSHFEIRAQPNDTVMAIKKNIEI
FT                   LIHNGKVLKDESTLAESKISEDGFLVVMLGKSKTMSSTGTPAAQSS
FT                   APAPAAAPASAVIPNTTVPEAPLSPAFAPSDTYGEAASNVVAGSNI
FT                   MWDTNMVSRALRAAYNNPERAVDYLYSGIPEMAEAAVPVSHFQGDQ
FT                   AGAAPGAPNSLPLNMFPQETLSGVTGAGLGSLEFLRNNPQFQTLRS
FT                   LELGKQNPQLLRQIQEHHEEFLQLINEPVEASEGDMFDQPEQDVPQ
FT                   RLEAMGFDRGLVIEAFLACDRNEELAVNYLLENAGDFED"
SQ   Sequence   379 BP;
>DCRAD23II_1
MKLTVKTLKGSHFEIRAQPNDTVMAIKKNIEDLQGKDNYPCGQQLLIHNGKVLKDESTLA
ESKISEDGFLVVMLGKSKTMSSTGTPAAQSSSAPAPTPAPAVAPAPAPAAAPASAVIPNT
TVPEAPLSPAFAPSDTYGEAASNVVAGSNLEQTIQHIMDMGGGMWDTNMVSRALRAAYNN
PERAVDYLYSGIPEMAEAAVPVSHFQGDQINAGNNAISDNGVAGAAPGAPNSLPLNMFPQ
ETLSGVTGAGLGSLEFLRNNPQFQTLRSMVQRNPQILQPMLLELGKQNPQLLRQIQEHHE
EFLQLINEPVEASEGDMFDQPEQDVPQEITVTAADQEAIERLEAMGFDRGLVIEAFLACD
RNEELAVNYLLENAGDFED
```

Alignment: 0964rad23.sid4 - pdb|1F4I|1F4I-A

BLASTP - alignment of 0964rad23.sid4 against pdb|1F4I|1F4I-A

uv excision repair protein protein rad23 homolog afragment: c-terminal uba domain;Mutant

- This hit is scoring at : 3e-05 (expectation value)
- Alignment length (overlap) : 45
- Identities : 46 %
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)
- Database searched : nrdb

```
Q:    324 EEQEAIGRLESMGFDRARVIEAFLACDRNEELAANYLLEHAGEED     368
          :E:EAI RL:::GF:.: VI:A:.AC::NE.LAAN:LL..  :::
H:      1 QEKEAIERLKALGFEESLVIQAYFACEKNENLAANFLLSQNFDDE      45
```

Legend of Alignment

```
:  positive score

.  score between -2 and 0
```

Please report problems and feedback concerning bioSCOUT through the support interface.

Alignment: 0964rad23.sid4 - pfam|hmm|ubiquitin

HMMPFAM - alignment of 0964rad23.sid4 against pfam|hmm|ubiquitin

Ubiquitin family

- This hit is scoring at : 91.1
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)

```
Q:       1 MKLTVKTLKGT-------HFEIRVQPNDTIMAVKKNIEEIQGkdsYPWGQQLLIFNGKVL
           MK: VKTL.G.         ..:.V:P:DT:::K:.IE: :G   .P  QQ LI: GKVL
H:       1 mkifVKtldgkgkvtdgstitlevepsdTVlelKekIedkeg...ippeqQrLiyaGkvL

            -KDESTLEENKVNEDGFLVVMLSKGK      78
             :D::TL.E .:.E..F.: :: : :
            reDdttLaeYniqegsftlhlvlrlr      83
```

Legend of Alignment

```
:  positive score

.  score between -2 and 0
```

Please report problems and feedback concerning bioSCOUT through the support interface.

Protein families database of alignments and HMMs

ubiquitin

Figure 1: 1d3z
Hydrolase
Ubiquitin nmr structure

Accession number: **PF00240**

Ubiquitin family

This family contains a number of ubiquitin-like proteins: SUMO (smt3 homologue) (see ULP1_YEAST), Nedd8 (see NED8_MOUSE), Elongin B (see Q15370), Rub1 (see Q9SHE7).

NEW! This family forms **structural complexes** with other Pfam families, to view them click here

---

INTERPRO description (entry IPR000626)

Ubiquitin [MEDLINE:91274342], PUB00000768, PUB00000768 is a protein of seventy six amino acid residues, found in alleukaryotic cells and whose sequence is extremely well conserved from protozoan to vertebrates. It is widely known as a post-translational tag used to signal a protein's hydrolytic destruction. Other functions for ubiquitin, depend on its differential internal isopeptide linkages. In addition, several ubiquitin-like proteins have been discovered from genome-sequencing efforts, other structural studies, and genetic screens. These new data show that proteins with the ubiquitin domain are adaptable, transposable genetic elements, which have been appended to other genes and utilized for many different cellular functions, depending on the ubiquitin-like protein's identity, subcellular location, and method of covalent attachment. The post-translational ligation of proteins to members of the ubiquitin superfamily can signal many different fates for the target protein [Larsen and Wang (2002) J. Proteome Res. 1, 411-419.]

Ubiquitin is a globular protein, the last four C-terminal residues (Leu-Arg-Gly-Gly) extending from the compact structure to form a 'tail' important for its function. The latter is mediated by the covalent conjugation of ubiquitin to target proteins, by an isopeptide linkage between the C-terminal glycine and the epsilon amino group of lysine residues in the target proteins.

In most species, there are many genes coding for ubiquitin. However they can be classified into two classes. The first class produces polyubiquitin molecules consisting of exact head to tail repeats of ubiquitin. The number of repeats is variable. In the majority of polyubiquitin precursors, there is a final amino-acid after the last repeat. The second class of genes produces precursor proteins consisting of a single copy of ubiquitin fused to a C-terminal extension protein (CEP). There are two types of CEP proteins and both seem to be ribosomal proteins. There are a number of proteins which are evolutionary related to ubiquitin, including ubiquitin-like proteins from baculoviruses; mammalian proteins GDX, FAU and RAD23-related proteins; human spliceosome associated protein 114, proteins BAT3 and CKAP1/TFCB, and ubiquitin-like proteins SMT3A, SMT3C and SMT3B; yeast proteins RAD23, DK2 and SMT3; and C. elegans SMT3 and ubl-1 proteins.

For additional annotation, see the PROSITE document PDOC00271 [Expasy | SRS-UK | SRS-USA]

Alignment	Domain organisati n
◉ Seed (95) ○ Full (1322) Format Coloured alignment Get alignment Further alignment options here Help relating to Pfam alignments here	◉ Seed (95) ○ Full (1322) As a Graphic / As a Tree Zoom 0.5 pixels/aa. ☐ Bootstrap tree View Graphic NIFAS Applet To find out about the NIFAS tree-viewer, click here
Species Distribution	**Phylogenetic tree**
NEW! View alignments & domain organisation by species Tree depth : Show all levels View Species Tree	◉ Seed (95) ○ Full (1322) Download tree ATV Applet The trees were generated using Quicktree To find out more about ATV phylogenetic tree-viewer click here

Database References	
PDB You can find out how to set up Rasmol here	1a5r ; 20; 95; PDB 2 Pfam Rasmol (unix) CATH-PDBSUM SCOP-UK
PRINTS	PR00348
PROSITE	PDOC00271 [Expasy | SRS-UK | SRS-USA]
HOMSTRAD	UBQ
PFAMB	PB013905 PB018771 PB024244 PB038815 PB039385 PB047048 PB073 PB074986 PB075222 PB075546 PB076014
SYSTERS	ubiquitin
PANDIT	ubiquitin

Literature References
1. Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes. **Whitby FG, Xia G, Pickart CM, Hill CP;** **J Biol Chem 1998;273:34983-34991.**
2. Structure determination of the small ubiquitin-related modifier SUMO-1. **Bayer P, Arndt A, Metzger S, Mahajan R, Melchior F, Jaenicke R,**

Pfam specific information	
Author of entry	Finn RD, Griffiths-Jones SR
Type definition	Domain
Alignment method of seed	Clustalw
Source of seed members	Prosite

HMMER build information		
	Pfam_ls [Download HMM]	**Pfam_fs** [Download HMM]
Gathering cutoff	25.5 25.5;	11.2 11.2

Becker J;
J Mol Biol 1998;280:275-286.

3.
Structure of tetraubiquitin shows how multiubiquitin chains can be formed.
Cook WJ, Jeffrey LC, Kasperek E, Pickart CM;
J Mol Biol 1994;236:601-609.

4.
Structure of ubiquitin refined at 1.8 A resolution.
Vijay-Kumar S, Bugg CE, Cook WJ;
J Mol Biol 1987;194:531-544.

Trusted cutoff	26.2 26.1;	11.2 11.2
Noise cutoff	25.2 25.2;	11.1 11.1
Build method of HMM	hmmbuild -F HMM_ls SEED hmmcalibrate --seed 0 HMM_ls	hmmbuild -f -F HMM_fs SEED hmmcalibrate --seed 0 HMM_fs

For help on making stable links to this page click here

Comments or questions on the site? Send a mail to **pfam@sanger.ac.uk**

Alignment: 0964rad23.sid4 - pfam|hmm|UBA

HMMPFAM - alignment of 0964rad23.sid4 against pfam|hmm|UBA

UBA/TS-N domain

- This hit is scoring at : 51.0
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)

```
Q:     144 NVDTIINQLMEMGGGSWDKDKVQRALRAAYNN--PERAVEYLYSG     186
             :. :.QL.EMGGG :.:::.::ALRA. NN   ERAVE:L..
H:       1 ideekvkqLreMGGGvfsreeakkALratnnnsDverAvewLleh      45
```

HMMPFAM - alignment of 0964rad23.sid4 against pfam|hmm|UBA

UBA/TS-N domain

- This hit is scoring at : 44.7
- Scoring matrix : BLOSUM62 (used to infer consensus pattern)

```
Q:     324 EEQEAIGRLESMG---FDRARVIEAFLACDRN--EELAANYLLEH     363
            ::E.: :L..MG   F.R... :A. A.:.N   E A..:LLEH
H:       1 ideekvkqLreMGGGvfsreeakkALratnnnsDverAvewLleh      45
```

Legend of Alignment

```
:  positive score

.  score between -2 and 0
```

Please report problems and feedback concerning bioSCOUT through the support interface.

Protein families database of alignments and HMMs

UBA

Figure 1: 1efu
C mplex (two elongation factors)
Elongation factor complex ef-tu/ef-ts from escherichia coli

Accession number: **PF00627**

UBA/TS-N domain

This small domain is composed of three alpha helices. This family includes the previously defined UBA and TS-N domains. The UBA-domain (ubiquitin associated domain) is a novel sequence motif found in several proteins having connections to ubiquitin and the ubiquitination pathway. The structure of the UBA domain consists of a compact three helix bundle [1]. This domain is found at the N terminus of EF-TS hence the name TS-N. The structure of EF-TS is known and this domain is implicated in its interaction with EF-TU [2]. The domain has been found in non EF-TS proteins such as alpha-NAC P70670 and MJ0280 Y280_METJA [1].

NEW! This family forms **structural complexes** with other Pfam families, to view them click here

---

INTERPRO description (entry IPR000449)

UBA domains are a commonly occurring sequence motif of approximately 45 amino acid residues that are found in diverse proteins involved in the ubiquitin/proteasome pathway, DNA excision-repair, and cell signaling via protein kinases [MEDLINE:97025177]. The human homologue of yeast Rad23A is one example of a nucleotide excision-repair protein that contains both an internal and a C-terminal UBA domain.

The solution structure of human Rad23A UBA(2) showed that the domain forms a compact three-helix bundle [MEDLINE:99061330]. Comparison of the structures of UBA(1) and UBA(2) reveals that both form very similar folds and have a conserved large hydrophobic surface patch which may be a common protein-interacting surface present in diverse UBA domains. Evidence that ubiquitin binds to UBA domains leads to the prediction that the hydrophobic surface patch of UBA domains interacts with the hydrophobic surface on the five-stranded β-sheet of ubiquitin [MEDLINE:22075341].

For additional annotation, see the PROSITE document PDOC50015 [Expasy | SRS-UK | SRS-USA]

Alignment	Domain organisation
◉ Seed (162) ○ Full (473)	◉ Seed (162) ○ Full (473) ○ Context (51)
Format Coloured alignment	**As a Graphic** **As a Tree**
Get alignment	Zoom 0.5 pixels/aa. ☐ Bootstrap tree
Further alignment options here	View Graphic NIFAS Applet
Help relating to Pfam alignments here	To find out about the NIFAS tree-viewer, click here

Species Distribution	Phylogenetic tree
NEW! View alignments & domain organisation by species Tree depth : Show all levels View Species Tree	Seed (162) Full (473) Download tree ATV Applet The trees were generated using Quicktree To find out more about ATV phylogenetic tree-viewer click here

Database References	
PDB You can find out how to set up Rasmol here	1aip C; 2; 42; PDB 2 Pfam Rasmol (unix) CATH-PDBSUM SCOP-UK
URL	http://www.isrec.isb-sib.ch/profile/isrec96/poster.html
PROSITE	PDOC50015 [Expasy | SRS-UK | SRS-USA]
PROSITE_PROFILE	PS50030
COGS	COG0264 COG1308
HOMSTRAD	EF_TS
PFAMB	PB005809 PB015819 PB016122 PB038705 PB073598 PB073788 PB074(
SYSTERS	UBA
PANDIT	UBA

Literature References

1. The structure of the Escherichia coli EF-Tu.EF-Ts complex at 2.5 A resolution.
Kawashima T, Berthet-Colominas C, Wulff M, Cusack S, Leberman R; Nature 1996;379:511-518.

2. **Alignment of TS-N domain is in figure 7c.**
Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell.
Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV; Genome Res 1999;9:608-628.

3. Structur of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr.

Pfam specific information	
Author of entry	Bateman A
Type definition	Domain
Alignment method of seed	Manual
Source of seed members	Bateman A

HMMER build information		
	Pfam_ls [Download HMM]	**Pfam_fs** [Download HMM]
Gathering cutoff	17.3 17.3;	19.0 19.0
Trusted cutoff	17.3 17.3;	19.0 19.0
Noise cutoff	17.2 17.2;	18.9 18.9
Build method of HMM	hmmbuild -F HMM_ls SEED hmmcalibrate --seed 0 HMM_ls	hmmbuild -f -F HMM_fs SEED hmmcalibrate --seed 0 HMM_fs

Dieckmann T, Withers-Ward ES, Jarosinski MA, Liu CF, Chen IS, Feigon J;
Nat Struct Biol 1998;5:1042-1047.

For help on making stable links to this page click here

Comments or questions on the site? Send a mail to **pfam@sanger.ac.uk**